

Molecular Modeling of Membrane Embedded Proteins

THÈSE N° 5695 (2013)

PRÉSENTÉE LE 31 MAI 2013
À LA FACULTÉ DES SCIENCES DE LA VIE
UNITÉ DU PROF. DAL PERARO
PROGRAMME DOCTORAL EN BIOTECHNOLOGIE ET GÉNIE BIOLOGIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Thomas LEMMIN

acceptée sur proposition du jury:

Prof. F. Naef, président du jury
Prof. M. Dal Peraro, directeur de thèse
Prof. M. Baaden, rapporteur
Prof. C. Chipot, rapporteur
Prof. U. Röthlisberger, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2013

Acknowledgements

I would like to sincerely thank Professor Matteo Dal Peraro, my PhD advisor, for his guidance, enthusiastic encouragement and constructive critiques throughout my Thesis research.

For creating and maintaining a friendly and stimulating scientific environment in the Laboratory for Biomolecular Modeling, I am grateful to all my lab mates. A special thanks goes to Dr. Matteo Degiacomi, with whom I spent nearly four years side-by-side.

It was a great pleasure to collaborate and learn from Professors William DeGrado, François Gisou van der Goot and Patrick Fraering and their researchers, Dr. Graham Clinthorne, Dr. Asvin Lakkaraju, Shixu Yan and Mitko Dimitrov. The advice and assistance given by Dr. Cinque Soto was greatly appreciated.

I am also grateful to the Life Sciences and EPFL IT staff members, for their valuable help setting up and managing the computational resources of our laboratory, in particular Sébastien Ferrara, Nicolas Michel, Dr. Jacques Menu and Benjamin Barras.

Last but not least, I thank my family, K and friends for their constant support and encouragement.

Summary

Over the past years, molecular modeling and simulation techniques have had a major impact on experimental life sciences. They are capable of providing accurate insight into microscopic mechanisms, which are usually difficult to investigate experimentally. Moreover, the integration of experimental data with molecular modeling appears to be a promising strategy to better understand complex biological processes.

In this thesis, molecular dynamics simulation has been used in combination with experimental data to investigate two transmembrane proteins: (i) the bacterial chemoreceptor PhoQ and (ii) the Amyloid Precursor Protein (APP).

(i) Bacterial two-component system PhoQ and bacterial membranes

Two-component systems (TCSs) are signaling complexes essential for bacterial survival and virulence. PhoQ is the histidine kinase chemoreceptor of the PhoQ-PhoP tandem machine that detects the concentration of cationic species at the inner membrane of Gram-negative bacteria. A full understanding of the PhoQ signal transduction mechanism is currently hindered by the lack of a complete atomistic structure. In this thesis project, the first structural model of the transmembrane (TM) portion of PhoQ from *Escherichia coli* was assembled, by using molecular simulations integrated with cross-linking disulfide scanning data. Its structural and dynamic features induce a concerted displacement of the TM helices at the periplasmic side, which modulates a rotation at the cytoplasmic end. This supports the idea that signal transduction is promoted through a combination of scissoring and rotational movements of the TM helices. Knowledge of this complex mechanism is essential in order to understand how the chemical stimuli sensed by the periplasmic sensor domain trigger, via the relay of the HAMP domain, the histidine auto-phosphorylation and kinase/phosphatase activity at the cytoplasmic end.

The PhoQ sensor domain lies in close proximity to the membrane. Interaction with anionic lipids, such as phosphatidylglycerols (PG) and cardiolipins (CL), are thought to play a key role in PhoQ activity. Present in bacterial and mitochondrial membranes, cardiolipins have a unique dimeric structure, which carries up to two charges, i.e. one per phosphate group, and under physiological conditions, can be unprotonated or singly protonated. Exhaustive models and characterization of cardiolipins are to-date scarce; therefore an *ab initio* parameterization of cardiolipin species for molecular simulation consistent with commonly used force fields is proposed here. Molecular dynamics (MD) simulations based on these models indicate a protonation-dependent lipid packing. A noteworthy interaction with solvating mono- and divalent cations is also observed. The proposed models will contribute to the biophysical and biochemical characterizations of bacterial and mitochondrial membranes and membrane-embedded proteins.

(ii) Structural and dynamic properties of the Amyloid Precursor Protein

The Amyloid Precursor Protein (APP) is a type I membrane glycoprotein present at the neuronal synapsis. The proteolytic cleavage of its C-terminal segment produces amyloid- β (A β) peptides of different lengths, the deposition of which is an early indicator of Alzheimer's disease (AD). Recently, the backbone structure of the APP transmembrane (TM) domain in detergent micelles was determined by nuclear magnetic resonance (NMR, independently by two different experimental groups). The TM conformations of these two structures are however markedly different. One is characterized by a highly kinked α -helix, whereas the other is mainly straight. Molecular dynamics simulations showed that the APP TM region is highly flexible and its secondary structure is influenced by the surrounding lipid environment. The size of the embedding detergent micelles strongly affects the conformation of the APP α -helix, with solvation being the main driving force for the development of a helical curvature. Once embedded in a membrane bilayer, APP systematically prefers a straight helical conformation. This is also confirmed when analyzing *in silico* the atomistic APP population observed in double electron-electron resonance (DEER) spectroscopy. In summary, the APP transmembrane domain is highly flexible due to the presence of glycine residues and can readily respond to the lipid environment, a property that might be critical for proteolytic processing by γ -secretase enzymes.

The presented thesis work clearly shows how molecular simulations and their interplay with available experimental input can help advance the understanding of the mechanism of complex biological systems and processes on a molecular scale. These results, in particular, go well beyond the current understanding of the functioning of two transmembrane proteins relevant for human health. Furthermore, the computational approaches and procedures developed in these projects will hopefully promote novel integrated strategies for investigating biological systems.

Keywords: molecular dynamics, molecular modeling, integrative modeling, antimicrobial peptides, two-component system, PhoQ, cardiolipin, membrane, membrane-protein, Alzheimer's, Amyloid Precursor Protein

Résumé

Au cours de ces dernières années, les techniques de modélisation *in silico* ont eu un impact majeur sur les sciences expérimentales. Ces techniques sont capables de fournir des informations précises sur les mécanismes moléculaires, qui sont généralement difficiles à mesurer expérimentalement. De plus, l'intégration des données expérimentales avec la modélisation moléculaire semble être une stratégie prometteuse pour mieux comprendre les processus biologiques complexes.

Dans le cadre de ce travail de thèse, des simulations de dynamique moléculaire ont été utilisées en association avec des données expérimentales pour étudier deux protéines transmembranaires: (i) le récepteur bactérien PhoQ et (ii) la protéine précurseur de l'amyloïde (APP).

(i) Le système à deux composants PhoQ et les membranes bactériennes

Les systèmes à deux composants (TCS) sont des protéines de signalisation essentielles pour la survie des bactéries et leur virulence. PhoQ est le récepteur à activité histidine kinase du tandem PhoP-PhoQ qui répond aux variations de concentrations de ligands cationiques à la surface de la membrane interne des bactéries Gram-négatives. La compréhension du mécanisme de transduction du signal de PhoQ est actuellement entravée par l'absence d'une structure atomique complète. Dans ce projet de thèse, le premier modèle structural du domaine transmembranaire (TM) de PhoQ de *Escherichia coli* était assemblé en combinant des simulations de dynamique moléculaire avec des données expérimentales de cross-link disulfure. Un déplacement concerté des hélices TM au niveau périplasmique induit une rotation des hélices à l'extrémité cytoplasmique. Ces observations supportent l'hypothèse d'une transduction du signal par une combinaison de cisaillement et de mouvements de rotation des hélices transmembranaires. Ce changement conformationnel complexe est essentiel pour la compréhension du mécanisme qui lie la détection des stimuli chimiques par le domaine périplasmique au déclenchement de l'activité kinase / phosphatase du domaine cytoplasmique de PhoQ.

Le domaine senseur de PhoQ se trouve à proximité immédiate de la surface de la membrane. Les interactions avec les lipides anioniques, tels que les phosphatidylglycerols (PG) et les cardiolipines (CL), sont supposées jouer un rôle clé dans l'activité de PhoQ. Les cardiolipines ont la particularité d'avoir une structure dimérique unique, qui leur permet de posséder jusqu'à deux charges négatives en fonction de l'état de protonation des groupes phosphates. Elles sont présentes dans la membrane des bactéries et des mitochondries. Les modèles et caractérisations des cardiolipines disponibles à ce jour sont limités. Une paramétrisation *ab initio* des différents états de protonation de la cardiolipine est proposée dans le cadre de ce travail. Les simulations de dynamique moléculaire basées sur ces modèles indiquent une organisation des lipides dépendante de leur état de protonation. Des interactions caractéristiques entre les cardiolipines et des cations mono et divalents ont

également été observées. Les modèles proposés contribueront à la caractérisation biophysique et biochimique des membranes bactériennes et mitochondriales, ainsi que des protéines membranaires.

(ii) Les propriétés structurales et dynamiques de la protéine précurseur de l'amyloïde

La protéine précurseur de l'amyloïde (APP) est une glycoprotéine membranaire de type I présente dans les synapses neuronales. Le clivage protéolytique de son segment C-terminal produit des peptides amyloïdes β ($A\beta$) de longueurs variables. Le dépôt de ceux-ci est un indicateur précoce de la maladie d'Alzheimer (AD). Récemment, deux structures du domaine transmembranaire (TM) de l'APP ont été déterminées par résonance magnétique nucléaire (RMN) dans des micelles. Les conformations TM de ces deux structures sont cependant très différentes. L'une est caractérisée par une hélice α très incurvée, tandis que l'autre est essentiellement droite. Les simulations de dynamique moléculaire ont montré que la région APP-TM est très flexible et sa structure secondaire est fortement influencée par son environnement. La taille des micelles englobantes affecte fortement la conformation de l'hélice α de l'APP, la solvatation étant la principale cause qui induit la courbure de l'hélice. Une fois incorporée dans une membrane, l'APP préfère systématiquement une conformation hélicoïdale droite. Ces résultats ont été également confirmés lors de l'analyse *in silico* de la population APP observée en spectroscopie double electron-electron resonance (DEER). En résumé, le domaine transmembranaire APP est très flexible grâce à la présence de plusieurs glycines et peut facilement s'adapter à son environnement. Cette propriété peut être critique lors du clivage protéolytique par les enzymes γ -sécrétase.

Ce travail de thèse montre clairement comment des simulations moléculaires combinées avec des données expérimentales peuvent aider à faire progresser la compréhension, à l'échelle moléculaire, de mécanismes et processus de systèmes biologiques complexes. En particulier, les résultats présentés permettent d'aller au-delà de la connaissance actuelle du fonctionnement de deux protéines transmembranaires impliquées dans des maladies humaines. En outre, les approches de calcul et les procédures développées dans ces projets pourraient promouvoir de nouvelles stratégies intégrées pour l'étude des systèmes biologiques.

Mots-clés: dynamique moléculaire, modélisation moléculaire, peptides antimicrobiens, système à deux composants, PhoQ, cardiolipine, membrane, protéines membranaires, maladie d'Alzheimer, protéine précurseur de l'amyloïde

Table of contents

ACKNOWLEDGEMENTS	III
SUMMARY	V
RESUME	VII
TABLE OF CONTENTS	1
LIST OF ABBREVIATIONS	5
INTRODUCTION	7
1.1 MEMBRANES AND MEMBRANE-EMBEDDED PROTEINS	7
1.2 MOLECULAR MODELING OF BIOLOGICAL SYSTEMS	9
1.3 OBJECTIVES OF THE THESIS	10
METHODS	13
2.1 MOLECULAR MECHANICS	13
2.1.1 BONDED INTERACTIONS	14
Bond stretching	14
Angle bending	14
Torsional terms	14
Improper torsions	14
2.1.2 NON-BONDED INTERACTIONS	15
Electrostatic interactions	15
Van der Waals interactions	16
2.2 CLASSICAL MOLECULAR DYNAMICS	17
2.2.1 SETTING UP AND RUNNING A SIMULATION	18
2.3 ENHANCED SAMPLING TECHNIQUES	19
2.4 REFERENCES	21
TWO-COMPONENT SYSTEM PHOPQ FROM <i>E. COLI</i>	23
3.1 INTRODUCTION	23
3.2 ASSEMBLY OF THE TRANSMEMBRANE DOMAIN OF <i>E. COLI</i> PHOQ HISTIDINE KINASE	26
3.2.1 INTRODUCTION	26
3.2.2 RESULTS	29
The tetrameric assembly of the TM domain reveals an Asn202 inter-helical lock	29
Connections between the TM bundle and neighboring domains	31
Asn202 is crucial for the solvation of the TM domain	32
Solvation triggers the conformational rearrangement of the TM domain	34
MOLECULAR MODELING OF MEMBRANE EMBEDDED PROTEINS	1

A coupled TM movement provides insight into the signal transduction mechanism	36
3.2.3 DISCUSSION	38
3.2.4 MATERIALS AND METHODS	41
3.2.5 SUPPORTING INFORMATION	42
3.2.6 REFERENCES	47
3.3 NEAR-ATOMISTIC MODEL OF THE COMPLETE PHOQ HISTIDINE KINASE	53
3.3.1 REFERENCES	54
3.4 MODELING OF CARDIOLIPIN SPECIES	55
3.4.1 INTRODUCTION	55
3.4.2 RESULTS	57
Biophysical characterization of the cardiolipin bilayer	57
Exploring the conformational space of the cardiolipin polar heads	61
Cardiolipin interactions with cations	62
3.4.3 DISCUSSION	67
3.4.4 MATERIALS AND METHODS	69
Structural models and parameterization of the cardiolipin polar heads	69
Molecular dynamics simulations	69
<i>A posteriori</i> validation of the cardiolipin parameterization	70
3.4.5 SUPPORTING INFORMATION	71
3.4.6 REFERENCES	73
3.5 LIPID BUILDER FOR MEMBRANE MODELS	77
3.5.1 REFERENCES	78
3.6 CONCLUSIONS AND PERSPECTIVES	79
 <u>AMYLOID PRECURSOR PROTEIN</u>	 81
4.1 THE FLEXIBILITY OF THE TRANSMEMBRANE DOMAIN OF THE AMYLOID PRECURSOR PROTEIN	81
4.1.1 INTRODUCTION	81
4.1.2 RESULTS	82
The APP-TM domain has a predominantly straight helical conformation in a membrane bilayer.	83
The APP-TM domain explores α -helical kinked conformations in detergent micelles.	85
The APP-TM segment responds to structural perturbations by exploiting its flexibility	87
4.1.3 DISCUSSION	91
4.1.4 MATERIAL AND METHODS	93
4.1.5 SUPPORTING MATERIAL	95
4.1.6 REFERENCES	100
4.2 ALZHEIMER'S DISEASE MUTATIONS IN THE AMYLOID PRECURSOR PROTEIN	103
4.2.1 INTRODUCTION	103
4.2.2 RESULTS	103
4.2.3 DISCUSSION	106
4.2.4 MATERIAL AND METHODS	107

4.2.5	REFERENCES	108
	<u>CONCLUSIONS AND PERSPECTIVES</u>	<u>109</u>
	<u>ANNEXE</u>	<u>113</u>

List of Abbreviations

AD	Alzheimer's disease
AMP	Antimicrobial peptides
A β	Amyloid- β
APP	Amyloid Precursor Protein
CDL	Unprotonated cardiolipin
CLP	Singly protonated cardiolipin
DEER	Double electron-electron resonance
DODE	Dodecane
DPC	Dodecyl Phospho Choline
EPR	Electron Paramagnetic Resonance
FAD	Familial Alzheimer's disease
LMPG	Lyso Myristoyl Phosphatidyl Glycerol
MD	Molecular Dynamics
MTSL	Methanethiosulfonate
NMR	Nuclear Magnetic Resonance
POPC	Palmitoyl Oleoly Phosphatidyl Choline
POPE	Palmitoyl Oleoly Phosphatidyl Ethanolamine
RMSD	Root Mean Square Deviation
SD	Sensor domain
TCS	Two-component system
TM	Transmembrane domain
WT	Wild Type

Introduction

This thesis deals with the computational modeling and simulations of integral membrane proteins at the atomistic scale.

1.1 Membranes and membrane-embedded proteins

The biological membrane constitutes an essential envelop for all living organisms, forming a selective barrier that prevents synthesized molecules from leaking out of the cell and unwanted molecules from diffusing into the cell. The plasma membrane defines the outer boundaries of the cell and allows it to regulate the composition of the cytoplasm. Inner compartments, called organelles, are also formed in the cell and carry out more specialized functions. Biological membranes are mainly composed of a complex and dynamic distribution of lipids, proteins and sugars (Figure 1).

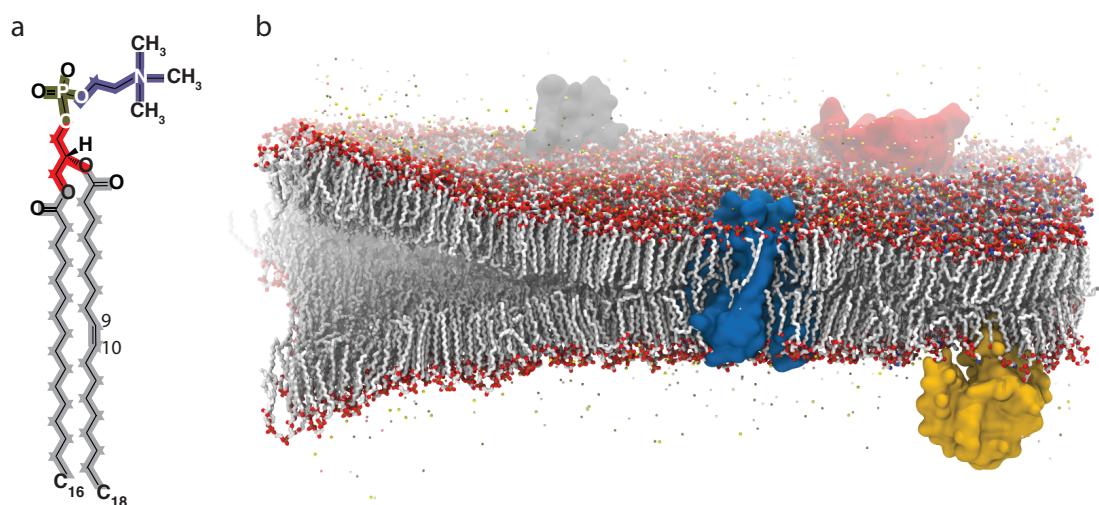


Figure 1: Biological membrane. (a) Phospholipid chemical structure: the fatty acids are highlighted in gray, the glycerol in red and the phosphocholine in gold and blue, and (b) structure of a biological membrane with embedded integral proteins.

The lipids composing the membrane are small amphipathic molecules that can be divided into three major species: phospholipids, glycolipids and cholesterol. Phospholipids are the predominant class in biological membranes and are composed of a hydrophilic head group and hydrophobic tails (Figure 1a). They can be divided into two classes: phosphoglycerides and sphingolipids. A glycerol group forms the

backbone of phosphoglycerides, which are attached to fatty acid chains and phosphorylated alcohol. They make up most of the plasma membrane lipids. Sphingolipids have a structure that is similar to phosphoglycerides, but contain a sphingoid base instead of the glycerol.

In contact with a polar solvent, amphipathic molecules usually arrange themselves into micelles to reduce the free energy. Phospholipids however are too bulky and therefore arrange themselves into bimolecular sheet-like structures mostly driven by hydrophobic interactions. These bilayers are composed of two leaflets, where the polar heads interact with the solvent, and the acyl chains form the hydrophobic core. The bilayers are typically 60 – 100 Å thick. Since they are energetically stable, disruptions and holes can readily be self-sealed, thus preserving the integrity of the cellular barrier (Figure 1b).

The chain length and degree of saturation of the acyl chains strongly affect the physicochemical properties of the membrane. The acyl chains are usually composed of an even number of carbon atoms, typically ranging from 14 – 24. The packing of the lipids strongly depends on the composition of the acyl chains. Longer or saturated acyl chains interact more favorably with each other. They therefore increase the melting temperature of the membrane and consequently reduce the membrane fluidity. For most biological membranes, the transition temperature is between 10 – 40° C. The presence of cholesterol in the eukaryotic membrane also influences its fluidity. The cholesterol fits between the phospholipids and stiffens the acyl chains, making the membrane less fluid.

The physicochemical properties of the different lipid species are therefore essential for the structural and functional organization of the biological membrane. They play a key role in the segregation of membrane domains, the formation of the non-bilayer structure, and the interaction with proteins. Several essential biological functions, such as signaling, mitosis and homeostasis rely on the specific lipid composition of the membrane and molecular interactions with other lipids, proteins and small molecules.

More than half of all proteins interact with the membrane. The protein content of the plasma membrane can be up to 50 %, or even 75 % for the outer membrane of the mitochondria. Depending on the function of the membrane, the lipid and embedded proteins differ. There are close interactions and modulations between the proteins and the membrane lipids. The major role of lipids however is to form a permeability barrier, whereas proteins have an important functional role. They guarantee the transport of molecules and information across the membrane. For example, membrane proteins permit the cell to sense its environment, to transport ions across the membrane or to produce energy.

Based on the nature of their interactions with the membrane, the membrane-proteins can be divided into two categories: the integral membrane-proteins (intrinsic) and the peripheral membrane-proteins (extrinsic). Intrinsic proteins extensively interact with the hydrocarbon chains of the membrane. One or more segments are embedded into the hydrophobic core, generally spanning the entire bilayer. These transmembrane domains are formed by one or more α -helices or multiple β -strands. The α -helical transmembrane domains compose the predominant category. The single pass

proteins can be classified as type I and type II transmembrane proteins, depending if the cytoplasmic tail is the C- or N-terminus, respectively. The β -strand transmembrane domains bind in an antiparallel arrangement, forming a hollow cylinder called a β -barrel. Thus far, they have only been found in bacteria and in some organelles such as mitochondria and chloroplasts. Certain membrane proteins do not enter the bilayer, but are anchored to a leaflet by a covalently linked hydrocarbon chain. An example is protein palmitoylation, where a palmitic acid is covalently attached to a cysteine or less frequently to a serine or threonine. The extrinsic proteins do not enter the hydrophobic core of the membrane. Instead, they bind electrostatically and by hydrogen-bond interactions to integral membrane proteins or with the head groups of the lipids.

Recent studies have highlighted how the biological membrane has a highly dynamic organization, where lipids and proteins can diffuse rapidly and are mutually dependent on one another. However, the fluid nature of the membrane coupled with the limited resolution of even state-of-the-art experimental techniques *in vitro* and particularly *in vivo* (e.g., high speed AFM), hinders a detailed molecular understanding of the fine interplay among lipids, small molecules, and proteins interacting at different levels with the membrane. Due to their location at the cellular envelope and their essential role in numerous physiological functions, the membrane proteins constitute preferential targets for drug development. An understanding of the dynamical and functional properties of membrane proteins, as well as the characteristics of the surrounding membrane is therefore crucial for improving or developing novel drugs.

1.2 Molecular modeling of biological systems

Earlier views of proteins and membranes as rigid structures have been progressively replaced by dynamic models in which internal motions and the resulting conformational changes play an essential role in their function. In recent years, molecular modeling and simulation have proven to be a valuable complement to conventional wet experiments to gain insight into biomolecular structure and dynamics. They provide an unmatched resolution able to reach the atomistic level. Moreover, the increase in computational power over the past two decades has had a great impact on extending the boundaries of molecular simulations. Originally limited to small proteins, at present they can be used to investigate large and heterogeneous biological systems at biologically relevant timescales.

A widely used molecular modeling technique applied to biological systems is Molecular Dynamics (MD). Given the structure of a biomolecule, all-atom MD simulations can be used to interpret experimental data and predict their functional properties. Processes can be modeled or reproduced at an atomistic level that carefully mimick physiological conditions, thus providing access to an accurate thermodynamic description of the system. The gain in computing power and MD code scalability permits studying larger systems (up to several million atoms), greater conformational changes and longer time scales (up to a μ s). At present, molecular dynamics is beginning to bridge simulated microscopic length and time scales to macroscopic experimental measurements. The application of MD simulations to biomolecular systems, integrated and guided by experimental data, has therefore

become an important and routine approach for better understanding the physical basis of emerging biological functions. In particular, the complexity of the membrane environment can now be finer modeled, allowing for a more detailed investigation of membrane-embedded proteins under physiological like conditions.

The accuracy of current force fields and the sampling of the phase-space are still two limitations in molecular simulations. First, parameterized models have to reproduce as closely as possible the real behavior of the considered system. Second, the accumulated sampling over biologically relevant timescales has to be carried out within a reasonable period of time. It appears that the precision of all-atom force fields is unexpectedly high and robust. Recipes have been developed to derive models for any cellular component. However, for many relevant systems, the time/length scale gap between computational and experimental methods is great and beyond the reach of current of all-atom MD simulations.

A possible strategy to overcome this limitation is to reduce the number of degrees of freedom in the system, which consequently decreases the algorithm complexity. Therefore, United-Atom (UA) and Coarse-Grained (CG) models have been developed, where fine intra- and intermolecular interactions are neglected in favor of a larger speed-up of simulations. When these approximations are not appropriate for the investigation of a biological system, enhanced sampling techniques can be coupled to MD simulations in order to explore more quickly the conformational space. In this case, the chemically-detailed representation is conserved, while the accessible timescales are largely amplified. Among these techniques, local elevation and metadynamics have shown to be valuable tools for simulating rare events in biomolecular systems, and thus reconstructing their associated free energy landscape.

1.3 Objectives of the thesis

At present, the experimental determination of high-resolution membrane protein structures is difficult. Combined with *in vitro* data, molecular modeling is therefore a valuable tool for the structural determination at near-atomistic resolution of this class of proteins. In this thesis, in close collaboration with several experimental groups, two transmembrane systems, linked to human pathogenesis, were mainly investigated using a dry/wet integrated approach: (i) the bacterial chemoreceptor PhoQ and (ii) the Amyloid Precursor Protein (APP) in eukaryotic cells.

Two-component system PhoQ from *E. coli*. Bacterial infections are one of the most frequent diseases in humans and animals, representing the second-leading cause of death worldwide. Basic research is therefore necessary to gain insight into the molecular function of bacteria. The first section of this thesis will focus on modeling the chemo-sensing protein PhoQ which is involved in the host recognition mechanism of bacteria. The PhoQ receptor plays a major role in the interaction of bacteria with their environment. It detects millimolar concentrations of divalent cations. The transduction of environmental changes across the biological membrane is essential for the survival of the organism. A modeling protocol based on a united-atom membrane mimic was used to assemble a model of the PhoQ transmembrane domain, and was cross-validated using experimental cross-linking data. The

implications of its structural and dynamic determinants are discussed to better rationalize the signal transduction mechanism. In a second phase, the parameterization of cardiolipin, a key anionic lipid species present at the bacterial membrane that plays an important role in PhoQ activity, was carried out. Cardiolipin microdomains were also assembled producing a model of the *E. coli* polar region. This more accurate bacterial membrane model will contribute to the investigation of membrane related processes, such as interactions with anti-microbial peptides, and lipid-mediated regulation of membrane proteins.

Amyloid Precursor Protein (APP). The Amyloid Precursor Protein is a type I transmembrane protein abundantly expressed in the human brain. A change of its proteolytic pathways plays a key role in the development of Alzheimer's disease (AD), the most common cause of dementia among elderly people. An exhaustive characterization of APP properties and interactions in the membrane are therefore required to advance the understanding of the molecular bases of Alzheimer's pathogenesis. Starting from recently reported NMR structures, molecular dynamics simulations were extensively carried out on the APP-TM to characterize its dynamics properties. APP-TM was inserted into diverse lipid environments. The effects of several different experimental setups were also investigated. The structural and dynamical features of each system were extracted and compared to the available Nuclear Magnetic Resonance (NMR) and Electron Paramagnetic Resonance (EPR) data. The MD simulations were crucial for the interpretation and rationalization of the experimental data.

This thesis is structured as follows: Chapter 2 presents the theoretical framework of the methods used for this work. Chapter 3 gives a description of the modeling and simulations performed for the investigation of the two-component system PhoQ from *E. coli*. and Chapter 4 details the work carried out on the Amyloid Precursor Protein. The conclusions and perspectives of the thesis are then presented and in the Annexe additional publications and studies related to this thesis are included.

Methods

Biologically relevant molecular systems are composed of thousands to millions of atoms. Such systems cannot be studied using first principles quantum mechanic calculations due to current theoretical and computational limitations. Therefore, classical molecular dynamics (MD) simulations are commonly used, because they are accurate enough to study the conformational space of biomolecules and are computationally efficient. They allow computing the time-dependent evolution of a molecular system at the atomistic scale. This chapter will present the theoretical framework of all-atom MD simulations, that were applied throughout this thesis project.

2.1 Molecular mechanics

Molecular mechanics is a classical mechanical technique for computing energies and selected properties of molecular systems. In an all-atomic molecular representation, the electronic motion is ignored and each atom is defined as a single particle. The potential energy of the system is calculated as a function of the nuclear position using a force field. Force fields are empirical models, designed to reproduce primarily the structural properties of molecules. They define the functional form and parameters describing the interactions between the atoms. Most atomic force fields have a similar functional form that is a compromise between accuracy and computational efficiency.

The basic molecular mechanics functional form describes the interactions between the atoms as a summation of bonded and non-bonded terms (eq. 1). The bonded terms are computed for the atoms covalently linked and the non-bonded terms are computed for all pairs of atoms that are at least separated by three bonds. The potential $U(r^N)$ depends on all the positions r of the N atoms composing the system (eq. 1),

$$U(r^N) = \underbrace{\sum_i^{bonds} U_{bond}(r_i) + \sum_i^{angles} U_{angle}(\theta_i) + \sum_i^{torsions} U_{torsion}(\phi_i)}_{bonded} + \underbrace{\sum_i^N \sum_{j \neq i}^N U_{coulomb}(r_{ij}) + U_{vdW}(r_{ij})}_{non-bonded} \quad (1)$$

The bonded terms are computed for all bonds, angles and torsional angles present in the molecule. The non-bonded terms model the electrostatic and van der Waals interactions between all pairs of atoms, i.e., i and j .

2.1.1 Bonded interactions

The bonded interactions are used to model all the intra-molecular bonded interactions. They account for the interaction of pairs to quartets of bonded atoms.

Bond stretching

The U_{bond} models the potential energy for the bond stretching between covalently bonded atoms. Since the forces between bonded atoms is very strong, it is rare that the distance between the atoms deviate significantly from its equilibrium position. In molecular mechanics, U_{bond} is often approximated by a harmonic oscillator (eq. 2),

$$U_{bond}(r) = \frac{k_i}{2}(r - r_0)^2 \quad (2)$$

In this case, the potential energy varies with the square of the displacement from the reference bond length r_0 (Hooke's law). Bond stretching is a strong interaction, *e.g.* a deviation of 0.2 Å from the reference in a C – C bond increases the system energy by ~12 kcal/mol.

Angle bending

The U_{angle} models the interaction between valence angles, *i.e.* three consecutive bonded atoms. It has also been described using Hooke's law (eq. 3),

$$U_{angle}(\theta) = \frac{k_i}{2}(\theta - \theta_0)^2 \quad (3)$$

Each angle is defined by the spring constant k_i and a reference angle θ_0 . Less energy is required for distorting than for the bond-stretching term.

Torsional terms

Since substantial energy is required to cause deformation for the bond-stretching and the angle-bending terms, most of the conformational changes in a molecule are due to the interactions between the torsional and non-bonded contributions. $U_{torsion}$ models the interaction between a bonded quartet of atoms. The torsional potential is almost always a periodic potential cosine series expansion (eq. 4),

$$U_{torsion}(\phi) = k_d(1 + \cos(n\phi - \phi_0)) \quad (4)$$

The value of k_d is a quantitative indication of the barrier height to rotation. The phase factor ϕ_0 defines where the energy minimum are. The multiplicity term n determines the number of minima.

Improper torsions

To accurately model the planarity and stereochemistry of chiral centers, an out-of-plane bending term $U_{improper}$ is often added to the force field (eq. 5),

$$U_{improper}(\phi) = \frac{k_i}{2}(\phi - \phi_0)^2 \quad (5)$$

The $U_{improper}$ considers the “improper” torsion angle composed of four consecutive bonded atoms not in their sequence. It is commonly modeled as a harmonic potential, where k_i represents the spring force constant and ϕ_0 the equilibrium angle.

2.1.2 Non-bonded interactions

The molecular structures are not only determined by the intra-molecular bonded interactions, but are also strongly dependent on the non-bonded forces between the atoms. In most force fields, the non-bonded terms are divided into two types: (i) the electrostatic and (ii) the van der Waals interactions.

Electrostatic interactions

An unequal distribution of charge in a molecule arises due to the difference in the electronegativity of its different constituting atoms. The electronegative atoms attract more the electrons and acquire a small negative charge. In an all-atom force field, each particle will have a specific partial atomic charge, designed to reproduce the quantum mechanic electrostatic properties of the molecules. Coulomb’s law is used to compute the electrostatic interaction as a sum of interactions between pairs of atoms (eq. 6),

$$U_{coulomb}(r) = \frac{q_i q_j}{4\pi\epsilon_0 r} \quad (6)$$

q_i and q_j are the partial charges of two atoms and ϵ_0 is the permittivity of free space. The electrostatic contribution is commonly ignored in the force fields of the bonded pair and triplet of atoms, since the stretching (1 – 2) and bending (1 – 3) terms already account for the interactions among atoms. For the atoms composing the torsional angle (1 – 4), the electrostatic interactions are either scaled down or ignored. The computation of electrostatic contributions between all the atoms of a system is expensive, *i.e.* it scales as N^2 , where N is the total number of atoms. Therefore, it is common to introduce a cut-off distance beyond which the electrostatic interactions are not calculated. However, for highly polar systems, this cut-off will lead to important errors due to the entry and exit of charged particles from the exclusion area. To reduce these errors, the Particle Mesh Ewald (PME) method is often used (1). Based on an Ewald summation, the electrostatic potential is separated into two terms: (i) a short-range potential, whose summation is efficiently carried out in the real space and (ii) a long-range potential carried out in the Fourier space. The use of the reciprocal space requires periodic symmetry for the system. Therefore the systems are required to be neutrally charged, typically by adding counter ions.

Van der Waals interactions

The Dutch scientist Johannes Diderik van der Waals observed and quantified deviations in the behavior of rare gas compared to ideal gas behavior. His findings suggested the presence of interactions different from electrostatic ones, since the multipole moments of these atoms are zero. The van der Waals interactions arise from a balance between attractive forces at long-range and repulsive forces at short distances and passes through a minimum at around $2.5 - 4 \text{ \AA}$ (e.g. 3.8 \AA for argon). These forces are due to two quantum effects: (i) attractive London forces and (ii) repulsive forces derived from the Pauli principle. The fluctuations in the electron clouds create instantaneous dipole in an atom, which in turn induces a dipole in the neighboring atoms. This creates an attractive induction effect, called the London or dispersive force. The Pauli principle prohibits two electrons in a system from having the same quantum number, resulting in a repulsive force (exchange force). This effect decays exponentially and explains the minimum observed for the van der Waals interactions.

The computation of these forces are complicated at the quantum level. Therefore in order to model this effect at the classic level, the best known van der Waals potential is the Lennard-Jones 12 – 6 function (eq. 7),

$$U_{LJ}(r) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - \left(\frac{\sigma_{ij}}{r} \right)^6 \right] \quad (7)$$

The Lennard-Jones potential is characterized by a single minimum of depth ϵ , and converges to zero at infinity. The parameter σ defines the collision diameter, *i.e.* the separation at which the energy is zero. The determination of the van der Waals parameters σ and ϵ for all pair of atom-types would be very time consuming. To overcome this problem, they are commonly obtained from the parameters of mono-atomic species using the Lorentz-Berthelot mixing rule. The collision diameter σ_{ii} is computed as the arithmetic mean of each atom (eq. 8). The well depth is equal to the geometric mean (eq. 9),

$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}) \quad (8)$$

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \quad (9)$$

To decrease the computational cost for the van der Waals interactions, a cut-off and a switching distance are commonly defined. The interactions of pairs of atoms beyond the cut-off distance are ignored. To avoid a discontinuity in the potential due to the cut-off, U_{LJ} is scaled at the switching distance so that it equals zero at the cut-off distance.

2.2 Classical Molecular dynamics

Based on molecular mechanics force fields, classical molecular dynamics (MD) aims at predicting the dynamical behavior of a molecular system. The successive conformations of the system are obtained by integrating Newton's second law (equation of motion) to determine the acceleration of each atom (eq. 10). This set of time-dependent configurations is called a trajectory.

Newton's equation of motion states that force \mathbf{F} applied to an ensemble of atoms at time t and position $\mathbf{X}(t)$ is equal to the negative gradient of the potential U experienced by the set of atoms,

$$\mathbf{F}(\mathbf{X}(t)) = -\nabla U(\mathbf{X}(t)) \quad (10)$$

The force on each atom will therefore change whenever the atom changes its position. To solve this equation, an iterative procedure is commonly used. The molecular dynamics trajectories are generated based on finite difference techniques, breaking down the integration into multiple steps separated by δt in time. The forces on each atom are determined at each time step based on its interactions with the other atoms. The acceleration is combined with the position and velocity of the atom at time t , allowing to calculate the new position and velocity at time $t + \delta t$. The forces are assumed to be constant during the time step.

Several integration algorithms have been proposed for molecular dynamics calculations and approximate the positions and dynamic properties of a molecular system as Taylor series expansions. At present, the Verlet integration algorithm is widely used for molecular dynamics simulations. The position of an atom at time $t + \delta t$ is computed based on its positions \mathbf{X} at times t and $t - \delta t$ and on its acceleration \mathbf{A} at time t (eq. 11). The Verlet integration algorithm does not explicitly consider the velocities, but they can be calculated from the derivative of the positions (eq. 12),

$$\mathbf{X}(t + \delta t) = 2\mathbf{X}(t) - \mathbf{X}(t - \delta t) + \delta t^2 \mathbf{A}(t) \quad (11)$$

$$\mathbf{V}(t) = \frac{[\mathbf{X}(t - \delta t) - \mathbf{X}(t + \delta t)]}{2\delta t} \quad (12)$$

As for any finite difference technique, a time step δt has to be defined for the Verlet algorithm. Unfortunately, there is no direct method to determine the correct time step. To guarantee the stability of a system, the Nyquist-Shannon sampling theorem states that the sampling frequency has to be at least two times faster than the fastest event of the system. In an all-atom molecular dynamics simulation, the C–H covalent bond vibrates with the highest frequency and has a repeat period of approximately 10 fs. A rule of thumb is to choose a time step ten times faster, and thus $\delta t = 1$ fs is commonly used for classical MD simulations. This time step can be increased to 2 fs,

by using an constraint algorithm, *e.g.* SHAKE (2) or LINCS (3) which constrain the bond length.

Since many experimental studies of biological molecules are carried out under conditions of constant temperature and pressure, a control of the temperature (300 – 310 K) and of the pressure of the system (~ 1 atmosphere) of the MD simulations is required.

The temperature can be directly related to the kinetic energy of the unconstrained system as follows:

$$\sum_{i=1}^N \frac{|p_i|^2}{2m_i} = \frac{k_B T}{2} 3N \quad (13)$$

where the moment p_i of N atoms is proportional to temperature scaled by the total number of degrees of freedom of the system. To control the temperature, thermostats are used to modify the velocities of the atoms, either by scaling them (Berendsen and Nose-Hoover thermostats) or via additional forces (Langevin),

In molecular dynamics simulation, the pressure of a system fluctuates much more than other quantities, such as the total energy or temperature. The virial theorem is used to calculate the internal pressure of the system (eq. 14).

$$P = \frac{1}{V} \left[Nk_B T - \frac{1}{3} \sum_{i=1}^N \sum_{j=i+1}^N r_{ij} f_{ij} \right] \quad (14)$$

The internal pressure of the system is proportional to the kinetic energy minus the virial term, *i.e.* the sum of the products of the coordinates r_{ij} of the atoms and the forces f_{ij} acting on them. The pressure control is achieved by using barostats that modify the positions of the atoms.

2.2.1 Setting up and running a simulation

In order to perform a molecular dynamics simulation, an initial configuration of the system has to be assembled. This is a key step for all MD simulations and is often critical for the success or failure of the simulations. Since the simulations are carried out at equilibrium, the initial configuration should be as close as possible to the equilibrium state. Therefore, the X-ray or NMR structures of the proteins are used as an initial configuration for biological systems.

Most MD simulations are carried out with periodic boundary conditions, to avoid boundary effects, since the system is of a finite size and use the PME algorithm calculate the electrostatic interaction. It is thus necessary to correctly set up the unit-

cell. First, it has to be big enough to avoid that the same atom interacts directly with a neighboring atom and its periodic image, but sufficiently small not to be too computationally demanding. For a biological system, a 12-15 Å (which is approximately similar to the short-range interaction cut-off, padding water box is commonly used to solvate a protein and defines the size of the unit-cell. In addition, MD simulations of the membrane protein require modeling the lipid environment. The protein is inserted into a membrane bilayer composed of the near-physiological mixture of different concentrations of phospholipids. Although most unit-cells are cubic, other geometries such as the hexagonal prism, truncated octahedron can be used. Second, the system is neutralized by adding counter ions, e.g. NaCl at near-physiological conditions (i.e., 150 mM).

All molecular dynamics simulations first start with a minimization of the system, in order to bring it to a local minimum of the potential energy surface. This is required to guarantee the stability of the system. Once minimized, it is necessary to assign initial velocities to the atoms, before starting the molecular dynamic simulation. The velocities can be randomly selected from a Maxwell-Boltzmann distribution at the temperature of interest (eq. 15).

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T} \right)^{1/2} \exp \left[-\frac{1}{2} \frac{m_i v_{ix}^2}{k_B T} \right] \quad (15)$$

This Gaussian distribution defines the probability p that an atom i of mass m has a velocity v in direction x . The initial velocities are often adjusted to ensure a zero momentum of the system. Having set up the initial configuration of the system and the corresponding velocities, the molecular dynamic simulation can be started. Since the diffusion of lipids is one order of magnitude higher than water, a longer equilibration is required for such systems.

Currently, molecular dynamics simulation are performed for systems composed of up to several millions of atoms. The computation can be carried out on standard workstations, for small systems (less than 10'000 atoms). However, for larger systems, a supercomputer with several tens of thousands of processor is required. The D. E. Shaw Research group has designed and built a massively parallel supercomputer specifically for molecular dynamics simulations (4).

2.3 Enhanced sampling techniques

At present, the computational power allows simulating large and complex biomolecules. However several phenomena of interest take place at time-scales of an order of magnitude higher than simulated time-scales and are not accessible using conventional molecular dynamics simulations. Enhanced sampling techniques such as local elevation (5), conformational flooding (6) or metadynamics (7) have proven to be essential to efficiently explore the conformational space of molecular dynamics simulations. These algorithms add a memory-dependant potential to the

simulation Hamiltonian, in order to force it to sample unexplored conformations of the system.

In this thesis project, metadynamics simulations were used to characterize the conformational space of membrane-embedded proteins. Metadynamics simulations were extensively used to study systems described at the classical (8,9) and quantum level (10,11). These simulations allow reconstructing the free energy landscape and accelerating rare events. This algorithm assumes that a small set of degrees of freedom, called *collective variables* are sufficient to describe the process of interest. The simplest type of collective variable is related to the geometry of the system, *e.g.* distances, angles or dihedrals. They can be applied to single atoms or to the center of mass of a group of atoms. Such variables are suitable to describe the conformational changes of biological systems, which are usually slow events. The reliability of metadynamics simulations depends on the correct choice of the collective variables. However, the optimal set is difficult, if not impossible, to determine *a priori*. For all practical applications, the collective variables have to satisfy at least the following three main properties: (i) Their value should significantly differ for the initial, intermediate and final states. (ii) The set should be small, *i.e.* two to three collective variables. Otherwise it becomes computationally too expensive to fill the free energy surface, and (iii) All the slow variables defining the phenomena of interest have to be included. Neglecting a relevant collective variable will lead to the overfilling a local minimum and the metadynamics simulation will never reach the free diffusion of the collective variable. Since no theoretical framework exists for choosing the correct collective variables, they rely on the chemical or physical intuition of the user and often have to be selected by a trial-and-error procedure. Several trial-and-error steps are usually required with different combinations of variables. A hysteretic behavior in the free energy always indicates that one or more relevant collective variables have been neglected. Metadynamics simulations have to be performed starting from a fully equilibrated molecular dynamics simulation. When setting-up the collective variables, it is useful to monitor them during the unbiased molecular dynamics, in order to gain important insight into their behavior. This will allow to better tune the collective variables during the metadynamics simulations.

Based on a set of collective variables $s(x)$, describing the phenomena of interest, a history-dependant potential is added to the classical MD Hamiltonian (eq. 16),

$$U_G(s(x), t) = w \sum_{t'=0}^t \exp \left(-\frac{\|s(x) - s(x(t'))\|^2}{2\delta s^2} \right) \quad (16)$$

During the metadynamics simulation, the position of the system in the collective variable space is determined and a Gaussian-shaped potential is added to the total potential. The added Gaussian potentials sum-up during the simulation and discourage the system to go back to already explored states until the system explores freely the collective variable conformation space. The free energy is finally recovered by the opposite sum of the added Gaussian. The metadynamics potential depends on three parameters: w and δs defining the height and width of the Gaussian and $t' = n\tau$, $n \in \mathbb{N}$. τ is the frequency at which the Gaussians are added.

Currently, no theoretical frame-work has been defined for choosing these parameters. As an initial guess for all-atom metadynamics simulations, Gaussians with a height w of 0.3 and a width δs of 0.4 are added every 300 steps. These values are then tuned after several trial-and-error steps. The width and height of the Gaussian determine the accuracy of the reconstructed free energy landscape. Increasing the height of the Gaussian is commonly used to quickly explore the conformation of the region of interest. This however leads to greater errors in computing of free-energy landscape. The error in the free energy landscape can be reduced by performing multiple metadynamics simulations and arithmetically averaging the resulting free-energy profiles.

In conclusion, metadynamics is a powerful tool for exploring the conformational space of a molecular system and for retrieving information about the corresponding free energy surface. However, it requires the user to choose (i) relevant collective variables, (ii) optimal metadynamics parameters, and (iii) to control the convergence and errors of the algorithm. The study of certain biological systems requires too many relevant degrees of freedom to allow the use of metadynamics. To overcome this problem, several new methodologies that combine a number of simultaneous replica of the system, were proposed, e.g. multiple walkers (12), parallel tempering metadynamics (13) and bias exchange (14).

In this thesis project metadynamics simulations were used to characterize the conformation space of the PhoQ (Chapter 3.2) and APP (Chapter 4.1) transmembrane domain.

2.4 References

1. Darden, T., York, D., and Pedersen, L. (1993) *J. Chem. Phys.* **98**, 10089
2. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. (1977) *Journal of Computational Physics* **23**, 327-341
3. Hess, B., Bekker, H., Berendsen, H. J., and Fraaije, J. G. (1997) *J. Comput. Chem.* **18**, 1463-1472
4. Shaw, D. E., Deneroff, M. M., Dror, R. O., Kuskin, J. S., Larson, R. H., Salmon, J. K., Young, C., Batson, B., Bowers, K. J., and Chao, J. C. (2007) Anton, a special-purpose machine for molecular dynamics simulation. in *ACM SIGARCH Computer Architecture News*, ACM
5. Huber, T., Torda, A. E., and Vangunsteren, W. F. (1994) *J. Comput. Aided Mol. Des.* **8**, 695-708
6. Grubmuller, H. (1995) *Phys Rev E* **52**, 2893-2906
7. Laio, A., and Parrinello, M. (2002) *Proceedings of the National Academy of Sciences of the United States of America* **99**, 12562-12566
8. Gervasio, F. L., Laio, A., and Parrinello, M. (2005) *J. Am. Chem. Soc.* **127**, 2600-2607
9. Biarnes, X., Bongarzone, S., Vargiu, A. V., Carloni, P., and Ruggerone, P. (2011) *J. Comput. Aided Mol. Des.* **25**, 395-402
10. Iannuzzi, M., Laio, A., and Parrinello, M. (2003) *Phys. Rev. Lett.* **90**

11. Ensing, B., Laio, A., Gervasio, F. L., Parrinello, M., and Klein, M. L. (2004) *J. Am. Chem. Soc.* **126**, 9492-9493
12. Raiteri, P., Laio, A., Gervasio, F. L., Micheletti, C., and Parrinello, M. (2006) *The Journal of Physical Chemistry B* **110**, 3533-3539
13. Bussi, G., Laio, A., and Parrinello, M. (2006) *Phys. Rev. Lett.* **96**, 90601
14. Piana, S., and Laio, A. (2007) *The Journal of Physical Chemistry B* **111**, 4553-4559

Two-component system PhoPQ from *E. coli*

3.1 Introduction

The increasing emergence of bacterial drug resistance is becoming one of the most significant challenges to human health. The effectiveness of the available weaponry against these pathogens is progressively lowered by the constant insurgence of single- or multi-drug resistant bacterial strains. Antimicrobial peptides (AMPs), which are produced by many tissues and cell types as part of their innate immune system, have attracted considerable interest as alternative antibiotics [1,2]. Unlike conventional antibiotics, AMPs operate via a non-specific mechanism. They target the bacterial cell membrane. This has the advantage of low bacterial resistance development [3]. Despite a large sequence diversity, AMPs are characterized by a cationic amphiphilic structure. They interact with the negatively charged membranes, disrupting the phospholipid membrane of the target cells. These types of antimicrobial peptides are known to mainly attack Gram-negative bacteria and only a few gram-positive bacteria [4].

Unfortunately, these peptides are complex molecules and suffer from proteolytic degradation. For a pharmacological use, it is necessary to capture the essential properties of the AMP as in an easy-to-synthesize molecule. However, a resistance to antimicrobial peptides is typically acquired by modification of the bacterial cell membrane, such as inserting lipopolysaccharides. Therefore, an optimal use of AMPs requires a complete understanding of the bacterial sensory systems and membrane. It has been proposed that the bacterial selectivity of AMPs is driven by the complementarity of their cationic nature with the anionic nature of the bacterial membrane. Thus, the bilayer composition is an essential determinant for AMP selectivity. The underlying mechanism is however not fully understood. A fine-tuned balance between several structural determinants in antimicrobial peptides is essential.

In bacteria, “two-component” systems (TCSs) detect the environmental changes via a sensor kinase at the periplasmic level, which leads to a phospho-relay cascade mediating specific gene transcriptions. They allow the bacteria to recognize and transduce the characteristics of extracellular environments [5]. Only a limited number of two-component proteins are found in a few eukaryotes, such as plants. They

appear to be absent in animals [6]. Therefore, this could be an ideal target for antimicrobial development.

Among the broad repertoire of TCSs, PhoPQ detects and responds to divalent cations, antimicrobial peptides [7] and low pH [8,9,10]. It is a major sensory system, regulating as much as 1% of the gene in certain Gram-negative bacteria [7]. PhoPQ belongs to a large family of membrane-associated protein kinases that undergo autophosphorylation and subsequently transfer the phosphate to the response regulatory domain (RR) (Figure 2). PhoPQ is composed of the histidine kinase PhoQ and the response regulator PhoP. *In vivo*, PhoQ forms a multidomain transmembrane homodimer, which molecular structure is largely unresolved. Consequently, the mechanism underlying PhoQ response to stimuli is not fully understood. A significant obstacle is the absence of structural information for the TM domain, which connects the structure of the periplasmic sensor domain to the cytosolic catalytic portion. Several mechanisms have been proposed to describe signaling transmission across the bacterial membrane, including piston shift, helix rotation and unwinding of a coiled coil [11,12]. Piston motion is thought to be the most efficient way to transmit the signal across the membrane [13,14], and is characterized by the displacement of the helix along its main axis perpendicular to the membrane's plane.

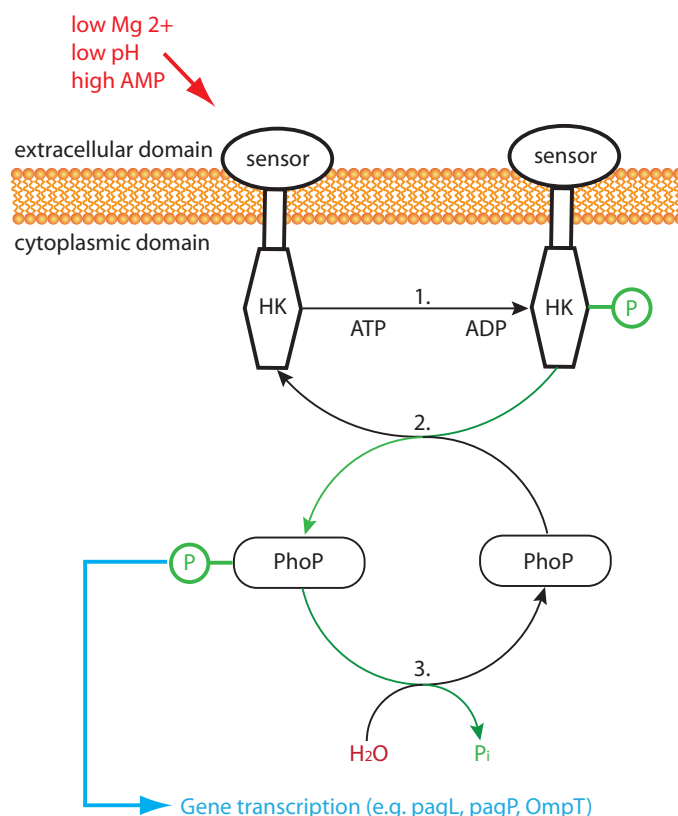


Figure 2: PhoPQ Phosphotransfer. For low concentration of Mg²⁺, acidic pH or in the presence of anti-microbial peptides, PhoQ is activated by phosphorylating H277 (1). The phosphoryl group is transferred to PhoP (2), which regulates gene expression. Finally PhoP is hydrolyzed, releasing the phosphoryl group (3).

Knowledge of these key membrane embedded systems is however not sufficient to understand the complex interplay between AMPs, and bacteria and their sensory systems. The lipid composition of the bacterial membrane is known to be essential for the function and modulation of the membrane protein. In the case of PhoQ, a large acidic cluster (148—EDDDDAE—154) at the sensor domain was originally proposed to be involved in the sensing of electrostatic variations at the bacterial membrane surface [15]. When the concentration of Mg^{2+} or Ca^{2+} cations is high (> 1 mM), cations can bind to the acidic cluster, forming a metal-mediated bridge with the negatively charged membrane of bacteria. Two major anionic constituents of the bacterial membrane are phosphatidylglycerols (PGs) and cardiolipins (CLs). It is noteworthy that cardiolipins are absent in mammalian cytoplasmic bilayer and are only found in the mitochondrial membrane. As such, cardiolipins could play a major role in AMP selectivity.

In bacteria, cardiolipins represent a small component ranging from 10 – 20% of the molar concentration and are not uniformly distributed in the membrane. The septal and pole regions of the bacteria are in fact enriched in CL up to 4-fold at the expense of PG moieties [16]. The bacterial membrane curvature is therefore influenced by the high local cardiolipin concentration. Furthermore, prokaryotes accumulate cardiolipins only under specific conditions, such as during stationary growth and environmental stress. The lack of CL reduces growth and function depending on oxidative processes. PG and CL deficient mutants are therefore not viable. In addition, cardiolipins act as a switch for protein interaction with cations, for example, as in the case of the EpsE system that is involved in the export of cholera toxin in *Vibrio cholerae* [17].

The first section of this chapter presents the modeling of the PhoQ TM domain. A novel scheme combining experimental cross-linking data and computational techniques was developed. The TM structural model allows the formulation of hypotheses concerning the transduction of the chemical signal and the computation of the corresponding free energy landscape. By combining the obtained TM model and homologous domains of the histidine kinase, an initial atomistic model of the PhoQ chemoreceptor will be proposed. The second section describes a novel parameterization of the cardiolipin, considering its different protonation states and its impact on the organization of the membrane. Interactions of mono and divalent cations were also investigated. A model of cardiolipin-rich domain characteristic of the polar region of *E. coli* is then presented.

3.2 Assembly of the transmembrane domain of *E. coli* PhoQ histidine kinase

Published as the following paper:

“Assembly of the transmembrane domain of *E. coli* PhoQ histidine kinase: implications for signal transduction from molecular simulations”

Lemmin T., Soto C. S., Clinthorne G., DeGrado W. F., and Dal Peraro M., PLoS Computational Biology, 2013

3.2.1 Introduction

Two-component systems (TCS) are protein signaling complexes present in most species of bacteria and are used to sense a wide range of environmental stimuli and couple them to adaptive responses [18]. The structure of a prototypical TCS consists of a membrane-spanning histidine kinase sensor, that senses the stimuli at the periplasmic region, and activates a cytoplasmic response regulator. The PhoQP TCS is reported to play a role in the defensive and virulence mechanism for certain Gram-negative bacteria [19,20]. External stimuli, such as the presence of antimicrobial peptides at the periplasmic surface, lead to the auto-phosphorylation of the PhoQ histidine kinase core, and to the subsequent transfer of the phosphoryl group to the response regulator, which elicits the regulatory response (kinase activity). The phosphorylated PhoP promotes the transcription of genes which leads to the modification of the outer membrane. These modifications increase the synthesis of enzymes that deacylate, almitoylate, palmitoylate, hydroxylate, and attach aminoarabinose to lipid A, thus promoting bacterial resistance to antimicrobial peptides (AMP) and reducing the host recognition of lipid A [21]. Only a limited number of two-component systems are found in a few eukaryotes [22]. Thus, apart from the basic understanding of the fundamental signaling mechanism, PhoQP TCS is a promising target for the development of synthetic antimicrobial drugs.

PhoQ forms a multidomain transmembrane homodimer, whose complete molecular structure is largely unresolved (Figure 1). Consequently, the mechanism underlying the PhoQ response to stimuli is not fully understood, and several mechanisms have been proposed to account for signaling transmission across the bacterial membrane, including piston shift, helix rotation and unwinding of a coiled-coil domain [11,12]. A major obstacle to elucidating the signal transduction mechanism of the PhoQP TCS is the lack of a full atomistic structure of PhoQ. The following four different regions can be defined and appear to have distinct functions (Figure 1): (i) The periplasmic region consists of a sensor domain (SD) that detects changes in the periplasmic environment. Two opposing crystal structures have been solved (1YAX [15] and 3BQ8 [23]); Goldberg *et al.* [24] have recently showed that the most probable physiological arrangement is described by the 3BQ8 x-ray structure. This is consistent with disulfide scanning experiments, and can explain the topology and connectivity of the SD with the remaining domains of PhoQ. (ii) Two membrane-

spanning antiparallel helices (called hereinafter TM1 and TM2) constitute the transmembrane (TM) region of PhoQ (Figure 3). (iii) Emerging from the cytoplasmic-facing membrane is a small signaling domain composed of a helix-loop-helix structure known as the HAMP region, and is thought to play an important role in transmitting the signal to the catalytic domain [25]. Finally, (iv) the catalytic or histidine kinase (HK) domain regulates the phosphoryl-state, and ultimately leads to PhoQ's function mainly as a kinase, or phosphatase during PhoP dephosphorylation. The phosphotransfer cycle between PhoQ and PhoP ensures, therefore, a robust and efficient genetic switch [26] (Figure 3).

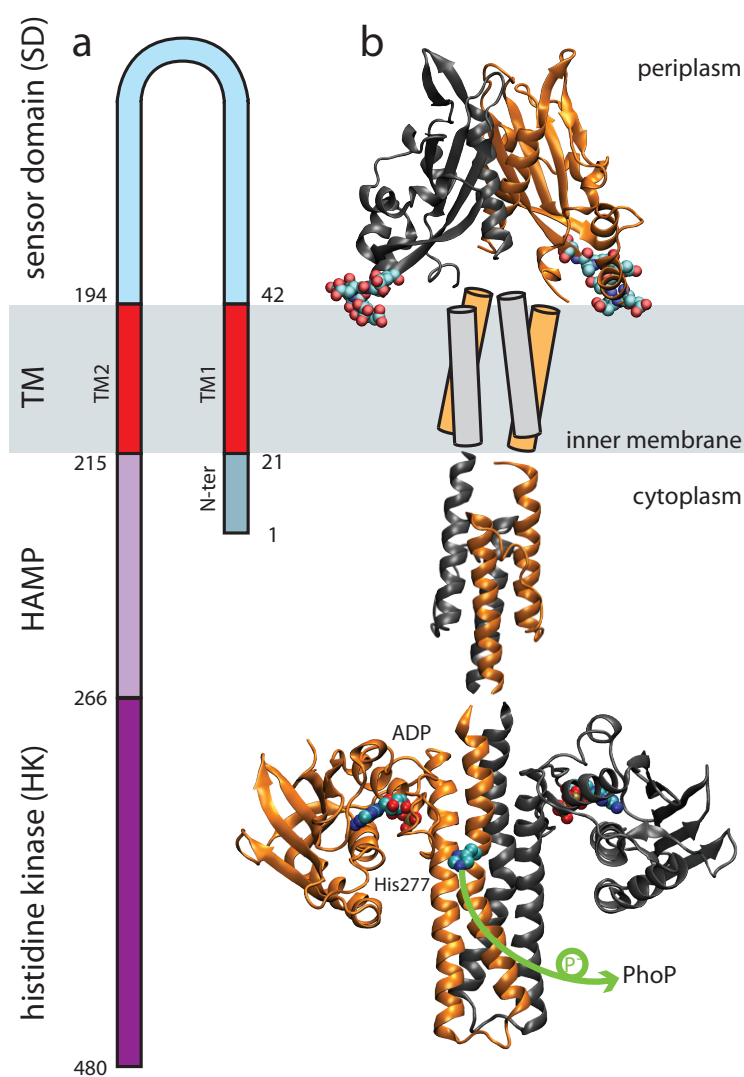


Figure 3: Schematic representation of the PhoQ histidine kinase in two-component systems. (a) The sequence topology of PhoQ is related to its (b) 3D homo-dimer schematic arrangement. The following available structures are used to model the global PhoQ structure: the sensor domain solved from *E. coli* (PDB code 3BQ8 [23]); the cytoplasmic HAMP NMR structure from *Archaeoglobus fulgidus* (PDB code 2L7H [27]), and the histidine kinase crystal structure from *Thermotoga maritima* (PDB code 2C2A [28]). The sensor domain harbors several acidic residues (148—EDDDAE—154, shown in space-filled representation). They

have been proposed to be involved in the sensing of divalent cations and AMP at the membrane surface. The ADP and the phosphorylation site (His277) are highlighted in space-filled representation. During the phosphor-transfer reaction, the phosphate group is transferred from His277 to the PhoP aspartic acid (Asp51) (green arrow).

In this study, we focus on the structural characterization of the transmembrane portion of PhoQ from *E. coli*. Difficulties in solving high-resolution x-ray structures of properly folded membrane proteins continue to be a significant barrier to determining structures at the atomic level and thus alternative, lower-resolution modeling strategies are often used as a proxy to obtain structural and functional information for membrane proteins [29,30,31,32,33]. Such strategies combined with computational methods can produce near-atomistic models; for example, replica exchange molecular dynamics (MD) was used to generate structural models for YycG sensor kinase consistent with mutagenesis studies [5]. The PhoQ protein might be a particularly good target for MD simulations since they are intrinsically well suited for studying the structures that are in dynamic equilibrium. When PhoQ activity is measured using transcriptional reporters, there is only a 2.5 to 3.5-fold difference in its activity in saturating Mg^{2+} concentration with respect to the basal state. This suggests a very small energetic difference in the two states, in the order of a single kcal/mol, assuming that the level of transcription reflects the fraction of the protein in the kinase mode. It is likely that multiple conformations observed by MD might be related to kinase or phosphatase activity. Thus, we used MD simulations to investigate the assembly of the PhoQ TM domain and the formation of a stable tetrameric bundle. MD simulations were guided and the resulting models validated by experimental cross-linking data. The TM domain connects with known, existing structures of the SD periplasmic domain as well as with the cytoplasmic HAMP domain. Therefore, the characterization of TM will provide insights into how conformational changes at the periplasmic SD might be transmitted to the cytoplasmic HAMP.

In order to test the validity of the models, we simulated point mutations involving the polar residue Asn202 [34], which appears to be relevant for the solvation and signaling function of the TM domain. PhoQ exists in equilibrium between phosphatase-active and kinase-active conformations, and mutations of this residue strongly influence the signaling ability of the protein. This probably supports a water-filled pocket, as also seen in TM domains of another HK structure, HTRII from bacteriorhodopsin [35]. We found that Asn202 is indeed crucial for the hydration of the PhoQ TM bundle. Importantly, this feature was found to directly trigger a combined scissoring movement and ~20 degree rotation of the TM helices that is propagated to the cytoplasmic side. This reveals the key role of the TM region in transmitting and modulating the signal through the bacterial inner membrane.

3.2.2 Results

The tetrameric assembly of the TM domain reveals an Asn202 inter-helical lock

Recently, disulfide cross-linking scanning was used to determine pairs of residues in close proximity in the PhoQ sensor domain [24]. The quantification of intermolecular disulfide formation showed a distinct periodic arrangement, fluctuating from almost 100% at peak efficiency to nearly zero at the lowest points, and guiding the modeling of the most biologically relevant sensor domain arrangement. The topology of the PhoQ homodimer indicates that four segments span the bacterial membrane (Figure 3). Therefore, by using the same procedure as for the sensor domain, disulfide cross-linking scanning experiments were performed for the TM region. Maxima occurred at residue positions 32, 35-36, 39-40 in TM1 and 199, 201-202, 205 in TM2, corresponding to the outer two-thirds of the TM domains (Figure S1 and Text S1). No significant cross-linking was observed near the cytoplasmic end of TM1, and the cross-linking was weak at this region of TM2, before gaining in intensity as the helix transitioned from the membrane and entered the cytoplasmic HAMP domain. The α -helical periodicity of the cross-linking results strongly suggested that the TM domain would assemble as a helical bundle, with the cross-linked residues facing towards the bundle core. These initial results along with available structural arrangements of domains connected to the TM (*i.e.* HAMP and sensor domain) strongly suggest a tetrameric helix assembly of the TM domain. We therefore modeled the TM1 and TM2 segments as ideal α -helices, and investigated their mutual interactions *ab initio* using MD simulations of the protein embedded in a lipid membrane bilayer; the disulfide cross-linking data was then used to validate the resulting models.

MD-refined models for TM1 (Thr21 to Val42) and TM2 (Phe195 to Trp215) showed only marginal fluctuations of the secondary structure (population RMSD = 1.4 ± 0.3 Å) (Figure 4a). This supports the predicted initial helical conformation and matches with the membrane hydrophobic core. The presence of a glycine (Gly33) in TM1 and a proline (Pro208) in TM2 are important characteristics of these helices. In fact, proline residues are known to induce distortions as large as 25 degrees with respect to the direction of the helix axis [36,37,38,39]; glycine may also create milder kinks in the helices [40,41]. Pro208 formed a kink in TM2 (28 ± 7 deg) as did Gly33 in TM1, but to a smaller extent (19 ± 5 deg) (Figure 4a). We expect that both kinks might favor the presence of a coiled coil-like helical assembly for the tetrameric TM1-TM2 complex. These equilibrated helices were used as initial atomistic models to study the assembly of the TM complex.

The homo- and hetero-dimer assembly of TM1 and TM2 helices were analyzed with MD simulations. In this case, the predictions of MD simulations can be directly compared to available disulfide scanning experiments, that have probed the mutual interactions of the TM1-TM1' and TM2-TM2' homo-dimers, and to a smaller extent the TM1-TM2 hetero-dimers, within the PhoQ TM complex. This set of simulations was conducted with a high concentration of TM segments in a united-atom lipid

bilayer to increase the conformational sampling (see Experimental procedures section). TM1 helices showed a strong tendency to form stable homo-dimers (60% of the total population, RMSD 2.1 ± 0.4 Å). In all TM1 dimers, hydrophilic residues (Thr21, Ser29, Tyr32, Tyr40) face the dimer interface (Figure 4b). Self-assembly of TM2 helices also produced a stable cluster of dimers (RMSD 2.7 ± 0.7 Å). The distortion caused by Pro208 led to a coiled coil-like assembly for TM2 dimers (Figure 4b). Asn202 plays a central role for the stability of the TM2 dimer. The Asn202 pair consistently forms two hydrogen bonds between their side chains. If only a single hydrogen bond is formed, then the dimer disassembles during the simulation. As previously reported [34], Asn202 is highly conserved in the transmembrane domain of TCS's and our models point to a structural role during dimer assembly. Moreover, experimental data indicate that the PhoQ function is impaired when Asn202 is mutated (e.g. during cross-linking), in agreement with this model, thus indicating a functional role of this residue in signaling conduction.

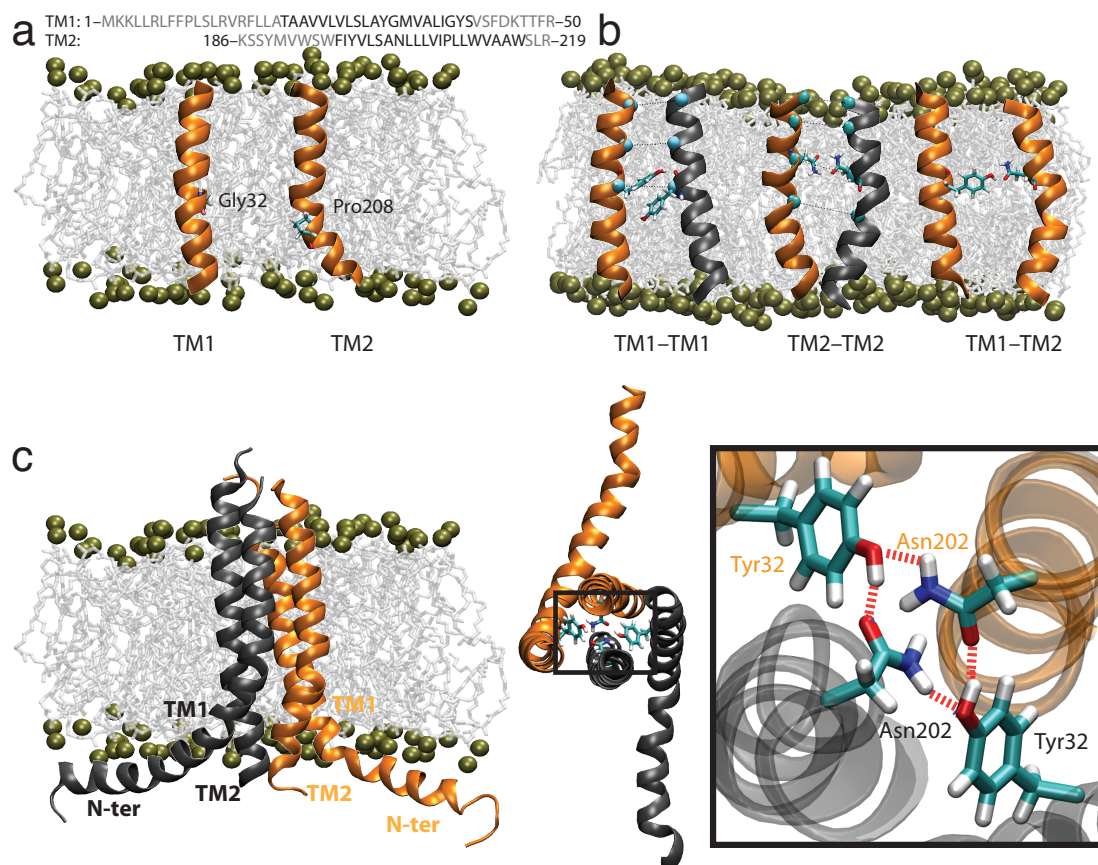


Figure 4: Assembly of the TM domain. (a) Models of individual TM1 and TM2 helices equilibrated in an all-atom membrane bilayer. Gly32 and Pro208 produce a kink in TM1 and TM2, respectively. (b) Stable TM dimers as obtained from MD in a united-atom membrane bilayer. (c) Tetrameric TM model structure produced using TM dimers, MD simulations and cross-linking spatial restraints. From left to right, the model is presented embedded in the membrane from the periplasmic top view, and with a focus on the H-bond network formed by residues Tyr32 and Asn202, which contributes to stabilize the bundle.

We also used experimental cross-linking data to validate the helix-helix interface of the dimer models produced by MD simulations. All residues showing high cross-linking efficiencies for TM1 and TM2 (TM1: Tyr32, Val35 Ala36, Tyr40; TM2: Leu199, Ala201 Asn202, Leu205) are in close proximity in the dimer models (Figure S2), thus providing a solid basis for further investigating the tetrameric assembly. We also probed the interactions along the TM1-TM2 interface: TM1 and TM2 domains associate in the membrane, forming oligomers. Two-thirds of the oligomers assemble in a conformation where Tyr32 interacts with Asn202 (Figure 4b). Moreover, these structures were in agreement with the cross-linking we were able to observe for the TM1-TM2 interface, where Ile207 was in close proximity to Val25 and Leu26 (Figure S3).

The TM1/TM2 homo/hetero-dimer structures were used to form a tetrameric conformation, and available cross-linking efficiency data were converted into semi-harmonic spatial restraints and applied during the first 10 ns of MD to equilibrate several initial tetrameric conditions. MD simulations were carried out afterward without any applied restraints on the bundle and converged to a stable conformation for the TM domain during ~80 ns of MD. The final structural ensemble of the TM domain is characterized by an approximate 2-fold symmetry (Figure 4c); TM2 helices are more tightly packed than to TM1 and this is consistent with the overall stronger cross-linking efficiency reported for TM2 (Figure S1). The coiled-coil TM1 and TM2 dimer is within a 1-Å RMSD of a Crick-ideal backbone [42,43]. All polar residues of TM1 (e.g. Thr21, Ser29, Tyr32, Tyr40) face the interface of the assembly. This final TM assembly is compatible with experimental cross-linking data, as shown by the correlation between the cross-linking efficiency and C_{α} contact maps calculated from MD simulations (Figure 5). The periodicity remains in close agreement with the predicted interface from cysteine-cross-linking mutagenesis. The predominant bundle conformation is stabilized by extensive van der Waals interactions throughout the bundle. Additionally, hydrogen-bonded interactions involving Asn202 and Tyr32 may contribute to stability. These polar amino acids formed a hydrogen-bonded network in many conformations, mainly involving the hydroxyl group from Tyr32 and the amide of Asn202 (Figure 4c). Previous studies showed that Tyr32 is not essential for signaling, and that various residues of different size, shape and hydrogen-bond donors and acceptors complementarity can replace Asn202 [34]. Thus, the specific hydrogen-bond network seen in the model may not be required for attaining the kinase-active conformation. Instead, it could be representative of structures formed in the resting phosphatase-active state.

Connections between the TM bundle and neighboring domains

The minimal TM structure was extended to help model the connections to the domains outside the lipid bilayer surface. N-terminus to TM1 is a 20-residue sequence with a high potential to form a surface-seeking amphiphilic helix [44], hereafter referred to as the cytoplasmic N-terminal amphiphilic helix (Figure 4c). At the C-terminus of TM1 we also included residues 43–50, corresponding to a portion of the helix that forms the dimer interface of the neighboring periplasmic sensor domain. We also extended the TM2 helix from its C-terminus to include its connection

to the HAMP domain. The additional segments were appended in a helical conformation to the TM structure, inserted in an all-atom membrane, and simulated over 150 ns of MD (population RMSD = 1.9 ± 0.3 Å). The TM bundle was stable during this timescale and maintained the same core interactions as described above (Figure 4c).

The N-terminal cytoplasmic helix folded onto the membrane in a surface orientation, as expected. The presence of a proline at position 10 induced a kink that bent the helix back making it parallel to the membrane surface in a rivet-like conformation (Figure 4c). This arrangement is also consistent with the amphipathic nature of the Met1 – Pro10 segment, and might be required for the anchoring and signal transduction through the membrane. The distance between TM2 C-termini is consistent with the NMR [27] and X-ray [45] structures of HAMP, and the periplasmic side of TM1 is well positioned to connect to the SD domain (namely pdb 3BQ8 [15]).

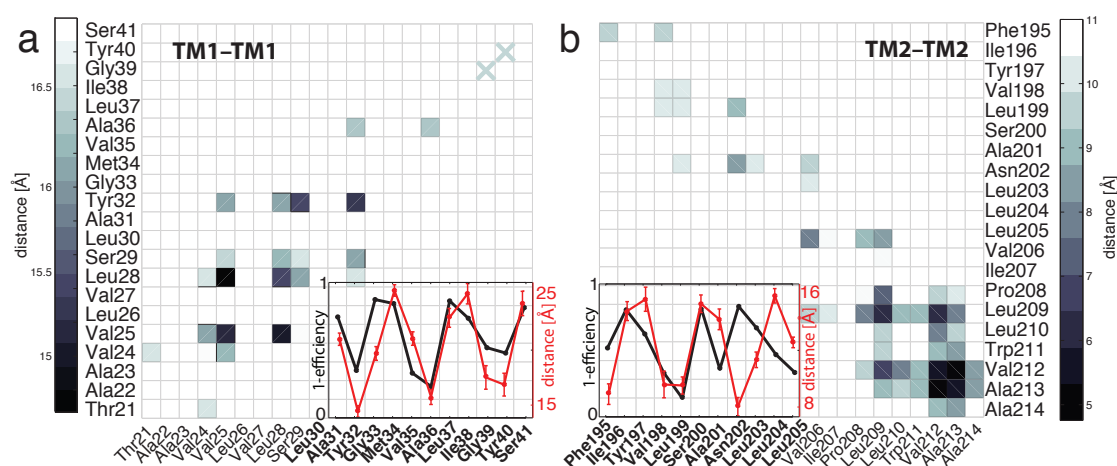


Figure 5: TM structural validation using disulfide cross-linking scanning. MD-averaged contact maps for (a) TM1 and (b) TM2 interfaces within the assembled TM domain. A direct comparison with cross-linking efficiency of (a) TM1 and (b) TM2 is reported in the inset, and shows a strong correlation between the cross-linking (1-efficiency) (black) and the MD-averaged C α distance measured for the TM model structure (red). The cross-linking efficiency for the whole TM1 and TM2 regions is reported in Figure S1.

Asn202 is crucial for the solvation of the TM domain

Recently, Goldberg *et al.* [34] showed that Asn202 in TM2 is critical for signal transduction. When Asn202 is mutated to non-polar residues, the transcription of PhoPQ-regulated genes is impaired. Thus, it was proposed that the polarity of Asn202 and the possibility to accommodate a water pocket at the TM core could potentially play an important role for kinase activity and signal transduction. Using the assembled model for TM, we further investigated the solvation of the bundle and monitored the structural determinants of Asn202 and other polar residues within the bundle (i.e. Tyr40, Ser43, Lys46, Thr47, Arg50, Lys186, Ser193, Tyr197, Ser200). Their arrangement forms a “hydrophilic ladder”, and allows water molecules to diffuse

into the membrane only from the periplasmic side. They progress discretely from residue-to-residue, by either interacting with the hydrophilic side chains or with the backbone. During MD, we observed transient solvation of the TM bundle. Lys46, Arg50 and Lys186, due to their long aliphatic chain and polar head, enhance water permeation in the membrane. Water molecules then access the membrane's hydrophilic core by interacting with Ser193, Thr47 or Ser43, before finally entering the bundle's center at the Asn202 position (Figure S4).

The presence of water in the middle of the bundle is consistent with the water-containing cavity hypothesis [34]. We introduced selected mutations in our model, namely N202A, N202R and N202H, to test their effect on the solvation of the bundle. *In vitro* experiments showed that N202A mutation fully impaired the kinase function of PhoQ. In MD, N202A resulted in an important rearrangement of the tetramer hydrogen bond network. The hydrogen bond network, observed for the wild-type TM bundle, is totally disrupted, and Tyr32 formed hydrogen bonds with the backbone of Ala202. Furthermore, the mutation prevented the water molecules from completely entering the bundle (Figure 6). *In vitro* experiments involving N202R and N202H showed increased kinase activity compared to the wild-type [34]. During the simulation, the N202R mutation, which was also the most "activating" mutation, had the greatest effect on hydration and structure. The Arg202 side chain stretched towards the periplasmic side of the membrane interface and attracted water molecules into the bundle (Figure 6). The core of the TM bundle remained solvated throughout the entire simulation. The snorkeling of Arg202 prevents Tyr32 from engaging in the same interactions as in the wild-type. The N202H mutant preserved the H-bond network: A single water molecule remained trapped in the hydrophilic cavity throughout the entire simulation (Figure 6). Our analysis seems to indicate that His202 is more efficient in keeping the cavity solvated. When the center of the bundle is solvated, TM1 bends towards the center of the tetramer for both mutations, reducing the inter-helical distance to approximately 17 Å. Thus, the electrostatic features and the general structure of our TM model allow for the solvation of the bundle. This fully supports the water-cavity hypothesis previously proposed and points to the crucial role of water for easing the mechanism involved in signal transmission defined by TM1 and TM2 segments (see following sections).

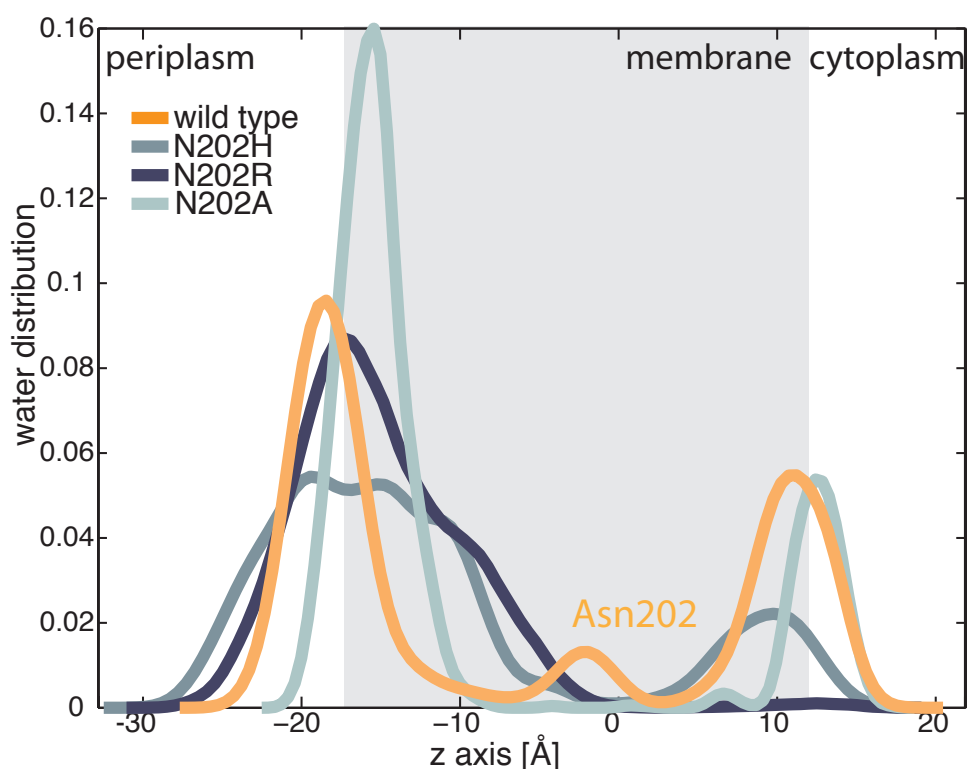


Figure 6: Effects of Asn202 mutation on the solvation of the TM domain. The kernel density estimation of water molecules for MD simulations of the wild-type TM bundle, and three relevant Asn202 mutants: N202A, N202H, and H202R. Residue 202 is localized in the middle of the membrane (at 0 Å). Conservative mutations preserve the hydration of the TM core, while substitution with alanine prevents water from entering the bundle. Distribution is calculated along the axis orthogonal to the membrane bilayer, and the transmembrane portion is schematically indicated by the grey area defined by the MD-averaged distance between bilayer polar heads (namely, phosphorus atoms).

Solvation triggers the conformational rearrangement of the TM domain

The transition between a solvated and non-solvated state observed in MD is associated with a conformational change in the wild-type bundle, i.e. a scissoring movement at the periplasmic surface of the membrane characterized by TM1 helices bending towards the center of the bundle, thus causing TM2 to spread apart. This scissoring movement, however, does not seem to propagate fully to the TM2 C-termini. To further characterize this conformational change, a principal component analysis (PCA) was performed based on the positions of the C_{α} atoms during the MD simulation. PCA has been shown to be an effective tool to remove thermal noise and retrieve significant movements of a system [46]. The main principal components were computed and used as a new coordinate system for the projection of the position of the TM bundle's C_{α} atoms. The rotation around the helix main axis was then extracted for TM2 (Figure 7a). The corresponding angle distribution has three modes that can be described using three Gaussians (Figure 7a).

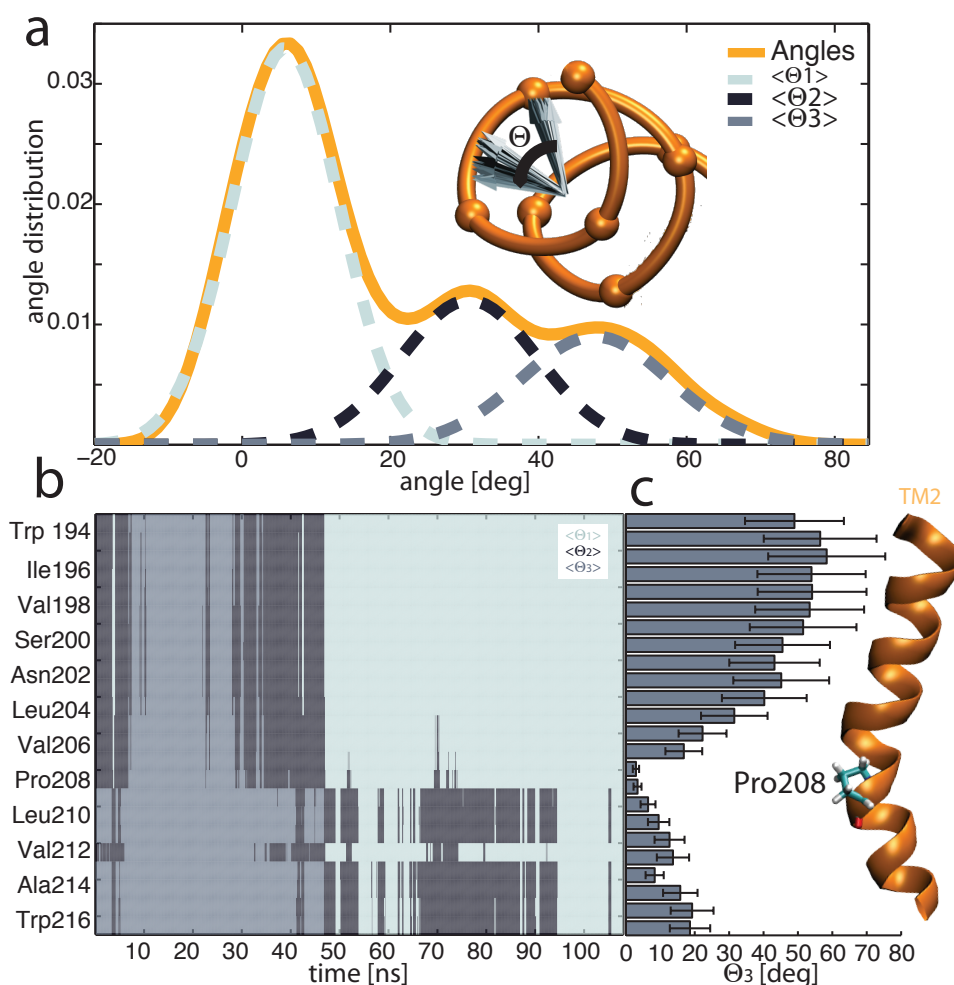


Figure 7: Solvation-dependent dynamic features of the TM domain. (a) The rotation of the C_α atoms around the TM2 helix main axis is computed based on a principal component analysis, and the angle distribution is characterized by three major modes that can be fitted using three Gaussians, (b) MD time series of the TM2 residue state corresponding to the angle distribution. The system is initially in a metastable state (black), before switching to a solvated state where the TM1-TM1 interface is tighter (grey). After ~30 ns, the system, passing through to a metastable state, shifts to a stable state characterized by a larger TM1-TM1 distance (light grey). (c) The rotation per residue related to the switch between the 2 most relevant states in MD (b) is calculated. TM2 Pro208 acts like a hinge, and transforms the large movement of the N-terminus into a mild rotation (~20 degrees) of the TM2 residue at the cytoplasmic interface.

Three centroids were defined by using the k-means clustering method, each representing a distinct state. In each frame of the simulation, the position of the C_α was classified in one of the three states. When plotted against time, each cluster corresponds to a well-defined conformational state in the simulation (Figure 7b). During the transition state, a minor (~10 deg) rotation of TM1 C-termini allows water molecules to enter the membrane and eventually solvate the center of the bundle. The solvation of this cavity triggers the previously described conformational change.

The scissoring movement of TM1 imposes a rotation of TM2 that is propagated to its C-terminus. It is interesting to note that Pro208 acts like a hinge, transforming a large TM2 rotation (> 45 degrees) coupled to a scissoring movement (6 \AA) at the periplasmic side into a pure TM2 rotation at the cytosolic side (Figure 7c). Cross-linking experiments showed that even though PhoQ was still functional, mutating Pro208 decreased the activity of PhoQ. This is consistent with sequence analysis using BLAST [47,48], where we found that proline at position 208 is highly conserved, thus supporting its functional role. The final rotation of TM2 C-termini is approximately 20 degrees. The helical conformation of the TM region also induces a minor coupled piston-like translation of TM2 ($0.8 \pm 0.1 \text{ \AA}$). This could influence the stability of HAMP and trigger the signal transduction [49]. Furthermore, the amplitude of the piston movement is in agreement with the experimental observation of the nitrate-sensing protein (NarX) structure [50]. This complex dynamic rearrangement was not observed for the N202A mutant that consistently remained locked in a rigid conformation with the lack of solvation of the bundle. Instead, the activating mutations (N202H and N202R) were able to explore the wild-type conformation characterized by a large solvation of the bundle, and are expected to readily adapt to conformational changes of the periplasmic SD to modulate the HAMP domain during PhoQ kinase activated state.

A coupled TM movement provides insight into the signal transduction mechanism

Unrestrained MD provided us with useful information about solvation and atomistic features of the bundle, including a possible coupled movement of TM1 and TM2. However, we were unable to exhaustively sample the large conformational space associated with TM dynamics, especially for higher energy states that could be relevant for the different kinase/phosphatase states of PhoQ. For this reason, we used enhanced sampling techniques to explore the dynamic determinants of the TM domain. Among those methods, metadynamics has been shown to be a valuable tool in simulating rare events and reconstructing the free energy landscape of biomolecules [51]. In order to characterize the free energy landscape associated with TM conformational changes, we used the distances between TM1 C-termini and TM2 N-termini as main collective variables. They appear to describe the major large-scale fluctuation of the TM domain in free MD. They allowed the sampling of the opening and closing movements taking place at the periplasmic side of the membrane. Metadynamics confirmed the equilibrium configuration ensemble (F_0) found in free MD characterized by an inter-helical distance of $\sim 20 \text{ \AA}$ for TM1 and $\sim 14 \text{ \AA}$ for TM2 (

Figure 8). It is interesting to note that the free energy landscape associated with TM1 and TM2 movement showed a broad valley ($\sim 5 \text{ kcal/mol}$ higher with respect to the F_0 equilibrium conformations), that extends to shorter TM1 ($\sim 16 \text{ \AA}$) and larger TM2 ($\sim 19 \text{ \AA}$) distances, thus capturing a second alternative conformation state (F_1), already transiently sampled during free MD simulations. Moreover, as in previous unbiased MD simulations, the conformational change along the valley is coupled with a ~ 20 degrees rotation of TM2 C-termini. This slightly higher energy region found in multiple metadynamics runs is probably representative of conformations sampled by TM

during signal propagation. In fact, given the small energy difference between the two states, which might be comparable or lower than the effect of changing the phospholipids composition of the bacterial membrane (event that is likely to happen during sensing of cationic species at the periplasmic side), we assume that both states might be representative of equilibrium-like configurations of either the kinase-active or phosphatase-active states.

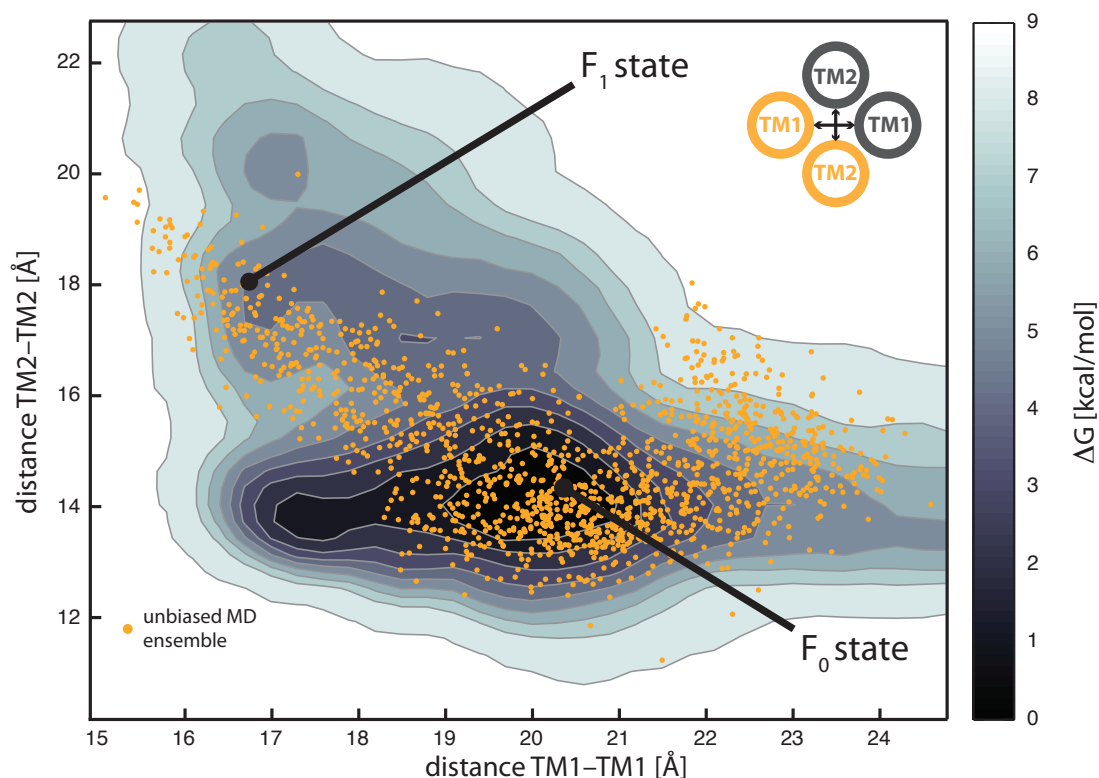


Figure 8: Free energy landscape for the TM conformational change. The free energy landscape defined by sampling inter-helical distances between TM1 C-termini and TM2 N-termini is reported. The conformational change observed in the unbiased MD simulations (orange points) occurs along a free energy valley, that connects a main equilibrium state (F_0) and a high-energy conformation, and can be associated with relevant states during the signaling process (F_1 , ~ 5 kcal/mol higher in free energy).

3.2.3 Discussion

Currently, no soluble NMR or X-ray crystal structure has been solved for the PhoQ transmembrane domain, thus hindering a thorough understanding of the signal transduction mechanism from the periplasmic sensor to the cytoplasmic kinase domain. In this study, we have combined all-atom molecular dynamics simulations with experimental cross-linking disulfide scanning data to build a four helical bundle model for the *E. coli* PhoQ TM region. Experimental cross-linking data provided a set of low-resolution spatial restraints for TM packing, thus validating a structural ensemble of the TM domain generated by MD simulations. The accuracy of the proposed structural ensemble might be limited by the absence of the adjacent PhoQ domains during the modeling and by the current sampling limitations of MD methods. However, the ensemble appears to recapitulate much of the experimental data observed for PhoQ. As discussed in more detail below, we see agreement with the experiments and find insights into the signaling of histidine kinases. First, the simulations provide an ensemble of molecular models that are consistent with the disulfide cross-linking data, particularly over the regions where strong cross-linking is observed. Secondly, they demonstrate the essential role of Pro208, which causes a bend in the TM2 helix. This bending provides a mechanism to modulate the amplitude and combination of motions in the TM domain. Finally, simulations of several mutants are compared to the experimental activity measurements, leading to a correlation between the degree of hydration of the bundle's core and the conformational changes of the TM domain.

Cross-linking data support a four-helix bundle for the PhoQ transmembrane domain (Figure S1). During MD, the packing of the side chains formed a stable coiled-coil tetrameric arrangement (within 1 Å RMSD of a Crick-ideal backbone), characterized by an approximate 2-fold symmetry. Moreover, we identify an N-terminal cytoplasmic amphiphilic helix lying along the membrane surface. This could help anchor TM to a specific location in the membrane bilayer and in turn, may favor the signaling transmission. Furthermore, the tetramer assembly is fully consistent with disulfide scanning data. The registration and packing of the helices in this structure is fully compatible with our model (RMSD 2.1 Å), and the structural difference between the two models is consistent with the resolution that is characteristic of the approach used to model the TM region. Although some aspects of the disulfide cross-linking studies are well explained by the model, others would appear more difficult to rationalize. For example, experimentally one observes very low cross-linking efficiency for TM1 and TM2 at the cytoplasmic interface (Figure S1). On one hand, the tight packing of this region, as observed in our model, could preclude the diffusion of reagents needed to induce disulfide formation. On the other hand, the diagonal helices of the TM bundle might also explore more distant conformations during signaling, as observed in the related TM bundle of HtrII, where the helices are under-packed in this region of the structure [35].

Despite this discrepancy, the topology of the bundle is consistent with the available structures of the periplasmic (SD) and cytoplasmic (HAMP) domains. At the periplasmic membrane surface, when the bundle is solvated, TM1 helices can easily connect to the biologically-validated dimeric conformation of the sensor domain (pdb: 3BQ8 [23]) (Figure 9). On the cytoplasmic surface of the membrane, the TM2 C-

terminal helices can be extended to merge with the two currently available structures of the HAMP domain (Figure 9). Furthermore, we observed that the kink induced by Pro10 at the PhoQ N-terminal amphiphilic helix is functional, thus helping to avoid steric clashes with the helical structure of the HAMP domain (Figure 5).

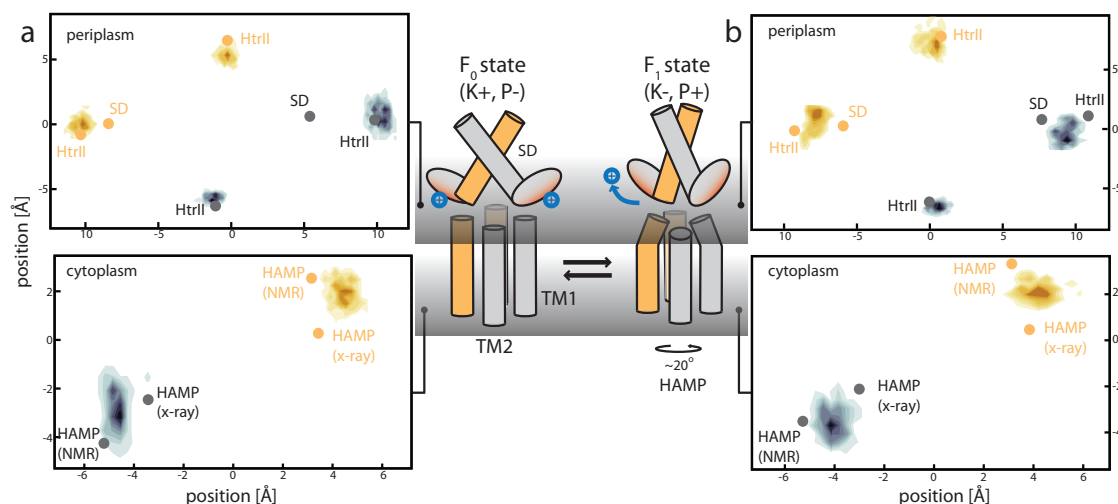


Figure 9: TM connectivity and implications for signaling mechanism. The panels (a) and (b) indicate the projection for the F₀ and F₁ states, respectively, on the periplasmic (top) and cytoplasmic (bottom) surface of the membrane of the helical termini of the TM model. The contour lines represent the position of TM1 and TM2 during MD, and the dots represent the position of SD and HAMP available structures at the same section surface. (Central panel) At high mM concentration of Mg²⁺ and Ca²⁺ cations, metal bridges are formed between the SD acidic cluster and the negatively charged membrane. This conformation can be associated to the F₀ state of our model and to a kinase-dominant state (K⁺). When the concentration decreases, the metal ion bridges are disrupted, leading to repulsion between the membrane and SD active site. This triggers a conformational change of the TM domain (F₁ state associated to a phosphatase-dominant conformation, P⁺), and results in a rotation of TM2 at the cytoplasmic interface, that is then transmitted to the linked HAMP domain.

MD simulations of this structural ensemble showed that the polar nature of Asn202 is crucial for the assembly and proper solvation of the TM domain. In the proposed model, Asn202 forms a hydrogen-bonded network with water and other residues in the lumen of the TM region. Furthermore, the proximity of other polar residues at the periplasmic end allows water molecules to enter the bundle and reach its core. We showed that the removal of this polar residue in the N202A mutant (for which *in vitro* the kinase activity is completely impaired [34]) prevents the solvation of the TM core, thus locking the domain into a rigid one-state conformation. Mutations to histidine or arginine partially preserve the electrostatic nature of the native Asn202. This produces a similar hydration of the TM core and indicates a more flexible conformation of the TM domain. Consistently, the kinase activity of these mutants is conserved and is even higher than the wild-type [34]. Therefore, the degree of solvation of the TM domain appears to be strongly correlated with the kinase activity

of PhoQ. This points to the key role of polar residues located within the bundle and water molecules in facilitating signaling through the membrane.

It is interesting to see how this model can help elucidate the mechanism of signaling, and how its structure and dynamics relate to the SD and HAMP domains. The SD X-ray structure (3BQ8 [23]) can optimally be connected to the TM state observed during MD where the TM1 helices are closer together (F_1 in

Figure 8). Therefore, the F_0 state, featuring the opening of TM1 helices, could be associated with the yet uncharacterized SD conformation. The stimuli (e.g. change in the concentration of cations) could drive the rearrangement of the SD dimer, causing it to pivot around its inter-helical dimeric interface (Figure 9). This scissoring movement would change the angle and inter-helical distance at which TM1 helices enter the membrane, thus promoting a concerted rearrangement of TM1 and TM2 helices, as observed in our simulations. Clearly, further studies are needed to unveil the dynamic features of the SD-TM coupling. Nonetheless, several experimental studies have shown that the TM helices undergo a structural rearrangement during signaling [52,53,54], and that a sensor domain-independent mechanism also exists [55]. Therefore, the complex interplay between scissoring and rotational movements at the periplasmic side of the membrane results, as revealed by our model, in a simple clockwise rotation of the TM2 helices at the cytoplasmic side (~20 degrees, Figure 7 and Figure 9). This is consistent with the cogwheel mechanism proposed for the HAMP domain [27] (Figure 9). Therefore, we conclude that the two relevant conformations observed in MD for the TM domain might be representative of the two main states of PhoQ, i.e. kinase and phosphatase active. And, the interconversion between these two states modulates the rotation of the cytoplasmic HAMP domain and, eventually, the phosphorylation and dephosphorylation ability of the histidine kinase domain (Figure 9).

Finally, this study shows how molecular simulations combined with low-resolution spatial restraints extracted from experimental cross-linking data can be used to investigate the structure and assembly of transmembrane proteins. Although full coarse-grained MD approaches can correctly assemble helix oligomers in membranes [56,57,58], we adopted here an atomistic representation to better describe the helical kinks and the important solvation effects at the PhoQ TM domain. This protocol is of general applicability and can easily be extended to study other transmembrane protein complexes. In this case, the characterization of the PhoQ TM domain, not only sheds new structural light on two-component system signal transduction across membranes, but also can be exploited for the design of specific drugs or peptidomimetics capable of impairing PhoQ assembly and function.

3.2.4 Materials and Methods

Structural models. No structural data exist for TM1 and TM2. Thus, we modeled their structures *ab initio*. We combined hieght topology prediction algorithms to isolate the transmembrane domain (TM) of PhoQ (Table S1). The transmembrane portion of TM1 was roughly identified from residue Thr21 to Val42, and TM2 from Phe195 to Trp215. Both transmembrane domains were modeled as ideal α -helices based on the combined results of secondary prediction tools such as HNN [59], Jpred [60,61], NetSurfP [62], PSIPred [63], ProteinPredict [64]. These initial predictions were used to build the atomistic models of TM1 and TM2 embedded in the lipid bilayer. The MD simulations were then used to equilibrate them and to find the correct match with the membrane hydrophobic environment.

Molecular dynamics simulations of the assembly. We used molecular dynamics (MD) simulations to further characterize the assembly of the TM domain. In a first step, the ideal helical TM models were each separately inserted and equilibrated in a 60 x 60 Å² Palmitoyl Oleoly Phosphatidyl Choline (POPC) membrane patch [65] to characterize their isolated structure in a phospholipid bilayer. Then, the self-assembly of TM1 and TM2 was studied separately, producing a high-concentration of TM proteins in a 100 x 100 Å² pre-equilibrated patch of united-atom dodecane membrane (DODE). The thickness of the DODE membrane is approximately equivalent the hydrophobic core of the POPC bilayer. Diffusion in an all-atom membrane model is slow and the protein oligomerization may require timescales not easily accessible within an all-atom MD. We inserted the equilibrated transmembrane domains TM1 and TM2 into a united-atom dodecane membrane (DODE) [66]. In a united-atom model, groups of atoms are clustered, thus decreasing the computation and diffusion time. In a second step, TM1 and TM2 dimer conformations compatible with cross-linking data were used to assemble a four-helix bundle. The resulting tetramer was inserted in a 55 x 55 Å² DODE membrane bilayer. Since an important rearrangement of the side chains is expected, harmonic restraints were also added to C α showing high cross-linking efficiency. Only TM2 dimers were restrained, due to the highly dynamic nature, i.e. Phe195, Val198, Asn202 and Val206 pairs were subjected to semi-harmonic restraints with force constant $k_f = 100 \text{ kcal}/\text{\AA}^2$ and the rest point x_0 to 12 Å. The harmonic restraints were removed after a 10 ns simulation. In addition, an extended version of the TM1-TM2 tetramer (TM1: Met1 – Arg50, TM2: LYS186 – ARG219), and three mutant species (N202A, N202H, N202R) were also tested in a 70 x 70 Å² Palmitoyl Oleoly Phosphatidyl Ethanolamine (POPE) membrane.

All simulations were performed using NAMD [67] engine with the CHARMM27 force field [68], including CMAP corrections. We used metadynamics to explore the free energy landscape associated with the opening and closing of the TM domain at the periplasmic side. In metadynamics, Gaussians are added to the energy surface and force the system to escape from local minima. This technique requires *a priori* knowledge of the degree of freedom relevant to a conformational change. In the current system, the observed conformational change was described by means of two collective variables representing the distance between the C-termini and N-termini of

TM1 and TM2 respectively, and was the most relevant motion from a principal component analysis of the unbiased MD simulations. The inter-helical distances were defined using the center of mass of residue Ser43 to Thr48 for TM1 and Ser188 to Trp192 for TM2. A set of three metadynamics simulations (~30 ns each) was performed using the collective variable module of NAMD. Gaussians of width 0.3 and weight 0.01 were inserted every 300 fs. The conformational space was sampled from 15 to 25 Å for TM1 and 10 to 20 Å for TM2 (Figure 8). The obtained free-energy profile was visualized as an iso-contour map with a grid spacing of 0.3 Å (Figure 8).

It should be mentioned that our models and simulations may be affected by several approximations: the assembly does not include the complete PhoQ homodimer, but only the TM domain and perturbations might be more evident at the protein termini; two collective variables may not be able to describe all the degrees of freedom of the system, leading to an approximation of the real energy landscape; furthermore, sampling can also be an issue to accurately describe solvation of the bundle. Therefore, multiple MD simulations were run from different initial conditions to ensure the reproducibility of the results. A list of all performed simulations is reported in Table 2.

Crosslinking reactions and analysis. Covalent chemical cross-linking has shown to be a useful technique to probe protein-protein interactions and provide low-resolution structural information of the protein-protein reciprocal arrangement [69] [24,31,70]. If both cysteine residues are close to each other (< 12 Å between two C α), the cross-linking efficiency will be high. The spatial interaction between the different amino acids can then be deduced and provides spatial restraints on the folding and packing of the PhoQ TM region. In the current study using a procedure similar as in [10], we report cross-linking results for the TM1-TM1 and TM2-TM2 intermolecular interactions for the TM tetrameric bundle (Figures 3 and S1) [71]. We also report a limited set of data for the inter TM1-TM2 interactions centered on Ile207 at TM2 (Figure S3). To crosslink cysteine residues in the transmembrane domain we used the oxidative catalyst copper (II) 1,10 phenanthroline (Cu(II)Phenanthroline), a small membrane permeable reagent that efficiently catalyzes disulfide bond formation in the membrane. We treated a 10 μ L sample of cell envelopes to 10 μ L of buffer containing 2mM or 0.2 mM Cu(II)Phenanthroline for a final concentration of either 1mM or 0.1 mM. Reactions were allowed to proceed for 30 minutes at room temperature, and then stopped by the addition of 20mM N-Ethyl Maleimide (NEM) and 20mM EDTA. Reactions were spun at 16'000 rpm at 4 degrees C to concentrate membranes. Suspension in 8M urea LDS loading buffer preceded gel loading. Western blotting was used to estimate cross-linking efficiency quantifying the fraction of cross-linked dimer to total visible protein (i.e. dimer/dimer + monomer, see details in Text S1 and Figure 5, Figure S1 and Figure S3).

3.2.5 Supporting Information

Molecular dynamics setup.

All simulations were performed using NAMD [67] engine, with the CHARMM27 force field [68], including CMAP corrections. TIP3P water [72] parameterization was used to describe the water molecules. The spatial overlapping of lipid molecules and protein were removed and the resulting protein-membrane system was solvated in a variable-sized water box, neutralized through the addition of NaCl at a concentration of 150 mM. The periodic electrostatic interactions were computed using particle-mesh Ewald (PME) summation with a grid spacing smaller than 1 Å. All systems were first minimized by 2000 conjugate gradient steps. POPE membranes were equilibrated for 500 ps at 300 K with a restraint on the proteins, water and ions. Since the DODE membrane was already equilibrated, this step was not required. All systems were subsequently gradually heated from 0 to 300 K in 800 ps with a constraint on the protein backbone scaffold. Finally, the systems were equilibrated for 3 ns at 300 K. Free molecular dynamics of all equilibrated systems were performed with a 2 fs integration time step using the RATTLE algorithm applied to all bonds. Constant temperature (300 K) was imposed by using Langevin dynamics [73], with a damping coefficient of 1.0 ps. A constant pressure of 1 atm was maintained with a Langevin piston dynamics [74], 200 fs decay period and 50 fs time constant.

Crosslinking reactions, Western blotting and analysis. Oxidized membranes were reconstituted in 20uL of loading buffer (Invitrogen LDS buffer plus 8M urea and 0.5M NEM) and heated for 10 minutes at 70 degrees C. 5uL of sample were loaded onto either a 7% or 3-8% gradient Tris Actetate gel (NuPage, Invitrogen). Proteins were separated by molecular weight with application of 150V for 60 minutes. Proteins were transferred to a nitrocellulose membrane by electrophoresis (iBlot, Invitrogen). Membranes were washed with Tbst (10mM tris pH 7.5, 2.5 mM EDTA, 50 mM NaCl, 0.1% Tween 20) and blocked with 3% BSA in Tbst. PhoQ* was probed using Qiagen penta-his antibody. His antibody was probed with HRP conjugated sheep anti mouse IgG. Proteins were visualized by exposure to ECL reagent (Amersham, GE health sciences) for 1 minute and exposure to film for 30 – 60 seconds. Pixel density histograms were generated using the ImageJ software, freely available from NIH (<http://rsbweb.nih.gov/ij/>), and crosslinking efficiency was determined using the fraction of cross-linked dimer to total visible protein (dimer/dimer+monomer).

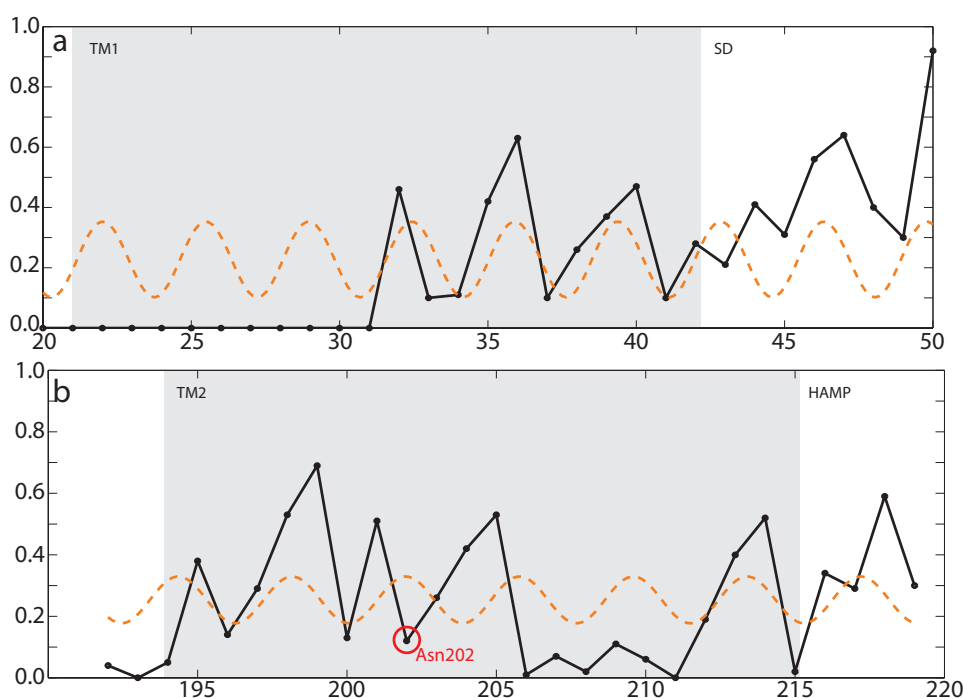


Figure S1: Fractional cross-linking of PhoQ TM residues. The black curve represents the cross-linking efficiency of (a) TM1 and (b) TM2. The data can be fitted with a sine wave (dashed orange). N202 was excluded for the fit, since it is an inactivating mutation. The reported periodicity ($\omega = 3.62$ and $\omega = 3.3$ respectively) strongly suggested that the TM domain would assemble as a helical bundle. These data are used to validate our results for TM assembly and are compared with MD results in **Figure 5** of the main text (see Supporting materials and methods section).

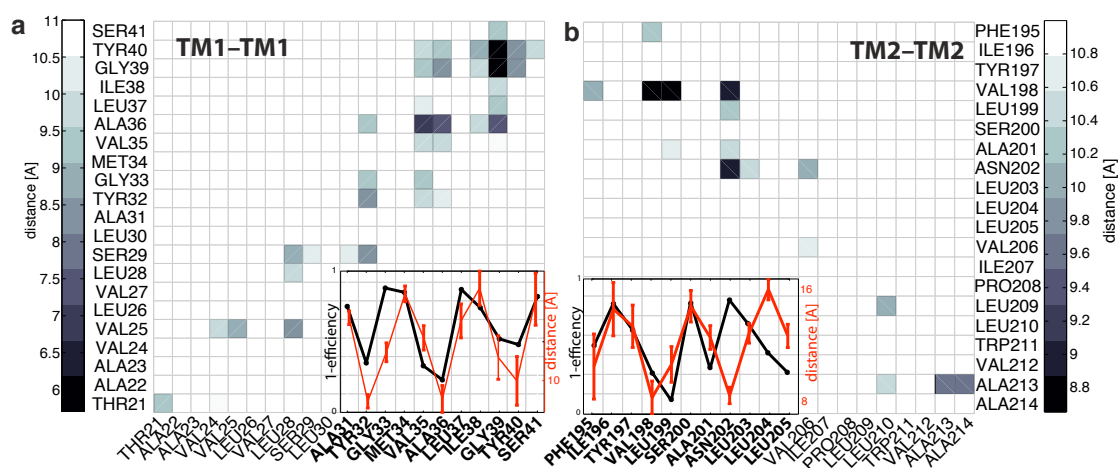


Figure S2: TM homodimer validation and selection via disulfide cross-linking scanning. MD-averaged contact maps for (a) TM1 and (b) TM2 dimer interfaces. A direct comparison with cross-linking efficiency (**Figure S1**) of (a) TM1 and (b) TM2 is reported in the insets and shows a strong correlation between the cross-linking (1-efficiency) (in black) and the MD-averaged C α distance measured for the TM model structure (in red).

Inter-TM crosslinking of Ile207

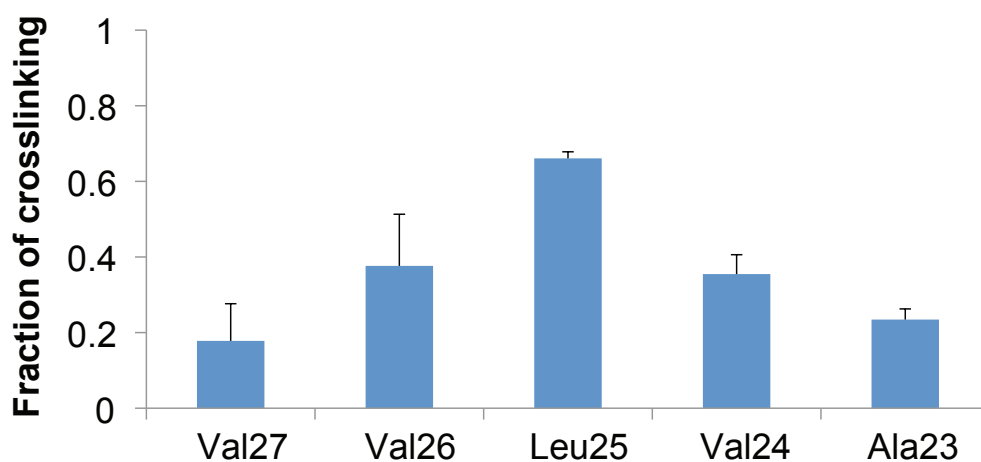


Figure S3: Inter-TM crosslinking for Ile207. Results of inter-TM crosslinking experiment between residue 207 on TM2 and a window of TM1 residues between 23 and 27. Error bars represent two independent experiments.

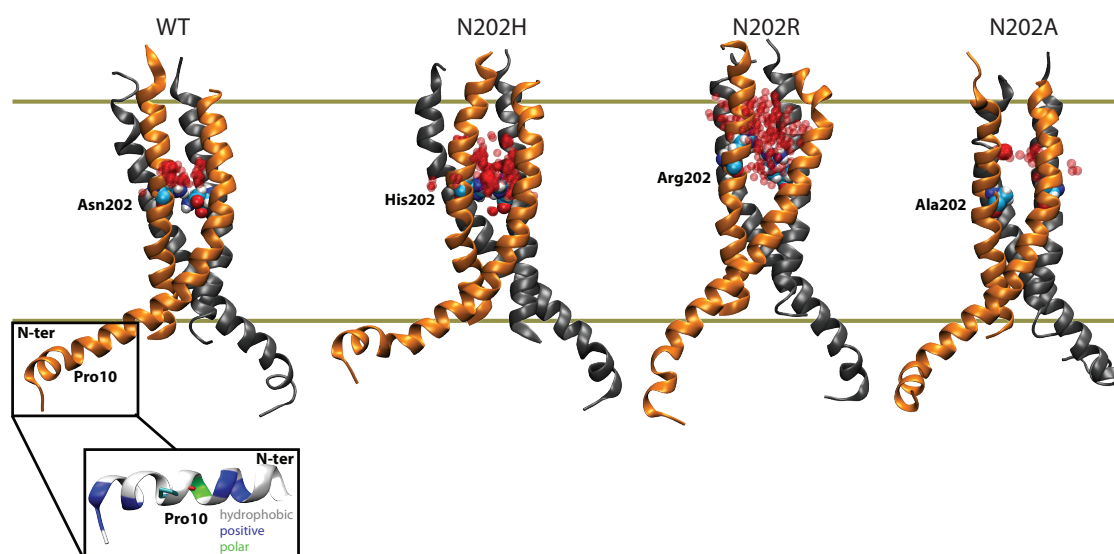


Figure S4: Water distribution around the TM bundle. (Top) The water molecules in contact with the TM bundle for wild type (WT), and three mutants: N202H, N202R and N202A are displayed with red spheres (see also Figure 4 in the main text). (Bottom) The focus on the N-terminal helix extended from TM1 illustrates its amphiphilic nature in our model. The hydrophobic residues are displayed in white, the polar and positively charged residue in green and blue, respectively.

Table 1: Transmembrane domain predictions. The transmembrane domain considered for the modeling is highlighted in grey as a consensus obtained from the following programs: SPOCTOPUS [75], OCTOPUS [76], PSipred [63], TMHMM [77], HMMTOP [78], TMpred [79], topcons single [80], and DAS [81].

	21	42	
	FLLATAAVVLVLSLAYGMVAL		uniprot
	FLLATAAVVLVLSLAYGMVALIG		HMMTOP
	ATAAVVLVLSLAYGMVALIGYSV		TMHMM
TM1	FLLATAAVVLVLSLAYGMVALIGY		TMpred
	FLLATAAVVLVLSLAYGMVALIG		Octopus
	FLLATAAVVLVLSLAYGMV		Soctopus
	FLLATAAVVLVLSLAYGMVALIGY		psired
	FLLATAAVVLVLSLAYGMVAL		topcons single
	FLLATAAVVLVLSLAYGMVAL		DAS
	195	215	
	FIYVLSANLLLVIPLLWVA		uniprot
	VWSWFIYVLSANLLLVIPLLWVA		HMMTOP
	WFIYVLSANLLLVIPLLWVA		TMHMM
TM2	FIYVLSANLLLVIPLLWVA		TMpred
	SWFIYVLSANLLLVIPLLWVA		Octopus
	SWFIYVLSANLLLVIPLLWVA		Soctopus
	FIYVLSANLLLVIPLLWVA		psipred
	FIYVLSANLLLVIPLLWVA		topcons single
	FIYVLSANLLLVIPLLWVA		DAS

Table 2: List of performed simulations

System	Membrane	Number of atoms	Simulation time
TM1 monomer	POPC	27,333	50 ns
TM2 monomer	POPC	27,956	50 ns
Dimerization TM1	DODE	40,700	100 ns
Dimerization TM2	DODE	40,700	100 ns
Dimerization TM1-TM2	DODE	40,700	100 ns
Initial tetramer	DODE	40,000	60 ns
Initial tetramer	POPE	47,300	80 ns
Extended tetramer	POPE	51,200	150 ns
N202A	POPE	51,200	100 ns
N202H	POPE	51,200	100 ns
N202R	POPE	51,200	100 ns
Metadynamics	POPE	51,200	3x30ns

3.2.6 References

1. Nusslein K, Arnt L, Rennie J, Owens C, Tew GN (2006) Broad-spectrum antibacterial activity by a novel abiogenic peptide mimic. *Microbiology-Sgm* 152: 1913-1918.
2. Toke O (2005) Antimicrobial peptides: New candidates in the fight against bacterial infections. *Biopolymers* 80: 717-735.
3. Tew GN, Clements D, Tang HZ, Arnt L, Scott RW (2006) Antimicrobial activity of an abiotic host defense peptide mimic. *Biochimica Et Biophysica Acta-Biomembranes* 1758: 1387-1392.
4. Hancock REW, Diamond G (2000) The role of cationic antimicrobial peptides in innate host defences. *Trends in Microbiology* 8: 402-410.
5. Szurmant H, Bu L, Brooks CL, 3rd, Hoch JA (2008) An essential sensor histidine kinase controlled by transmembrane helix interactions with its auxiliary proteins. *Proc Natl Acad Sci U S A* 105: 5891-5896.
6. Gao R, Stock AM (2009) Biological insights from structures of two-component proteins. *Annu Rev Microbiol* 63: 133-154.
7. Groisman EA (2001) The pleiotropic two-component regulatory system PhoP-PhoQ. *Journal of Bacteriology* 183: 1835-1842.
8. Gunn JS, Richards SM (2007) Recognition and integration of multiple environmental signals by the bacterial sensor kinase PhoQ. *Cell Host & Microbe* 1: 163-165.
9. Prost LR, Daley ME, Bader MW, Klevit RE, Miller SI (2008) The PhoQ histidine kinases of *Salmonella* and *Pseudomonas* spp. are structurally and functionally different: evidence that pH and antimicrobial peptide sensing contribute to mammalian pathogenesis. *Molecular Microbiology* 69: 503-519.
10. Prost LR, Daley ME, Le Sage V, Bader MW, Le Moual H, et al. (2007) Activation of the bacterial sensor kinase PhoQ by acidic pH. *Molecular Cell* 26: 165-174.
11. Gordeliy VI, Labahn J, Moukhametzianov R, Efremov R, Granzin J, et al. (2002) Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex. *Nature* 419: 484-487.
12. Ottemann KM, Xiao W, Shin YK, Koshland DE, Jr. (1999) A piston model for transmembrane signaling of the aspartate receptor. *Science* 285: 1751-1754.
13. Lynch BA, Koshland DE, Jr. (1992) The fifth Datta Lecture. Structural similarities between the aspartate receptor of bacterial chemotaxis and the trp repressor of *E. coli*. Implications for transmembrane signaling. *FEBS Lett* 307: 3-9.
14. Falke JJ, Bass RB, Butler SL, Chervitz SA, Danielson MA (1997) The two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annu Rev Cell Dev Biol* 13: 457-512.

15. Cho US, Bader MW, Amaya MF, Daley ME, Klevit RE, et al. (2006) Metal bridges between the PhoQ sensor domain and the membrane regulate transmembrane signaling. *J Mol Biol* 356: 1193-1206.
16. Kawai F, Shoda M, Harashima R, Sadaie Y, Hara H, et al. (2004) Cardiolipin domains in *Bacillus subtilis* Marburg membranes. *Journal of Bacteriology* 186: 1475-1483.
17. Camberg JL, Johnson TL, Patrick M, Abendroth J, Hol WG, et al. (2007) Synergistic stimulation of EpsE ATP hydrolysis by EpsL and acidic phospholipids. *EMBO J* 26: 19-27.
18. Bourret RB, Silversmith RE (2010) Two-component signal transduction. *Curr Opin Microbiol* 13: 113-115.
19. Hancock RE, McPhee JB (2005) *Salmonella*'s sensor for host defense molecules. *Cell* 122: 320-322.
20. Miller SI, Kukral AM, Mekalanos JJ (1989) A two-component regulatory system (phoP phoQ) controls *Salmonella typhimurium* virulence. *Proc Natl Acad Sci U S A* 86: 5054-5058.
21. Kawasaki K, Ernst RK, Miller SI (2005) Inhibition of *Salmonella enterica* serovar Typhimurium lipopolysaccharide deacylation by aminoarabinose membrane modification. *J Bacteriol* 187: 2448-2457.
22. Wuichet K, Cantwell BJ, Zhulin IB (2010) Evolution and phyletic distribution of two-component signal transduction systems. *Curr Opin Microbiol* 13: 219-225.
23. Cheung J, Bingman CA, Reyngold M, Hendrickson WA, Waldburger CD (2008) Crystal structure of a functional dimer of the PhoQ sensor domain. *J Biol Chem* 283: 13762-13770.
24. Goldberg SD, Soto CS, Waldburger CD, Degrado WF (2008) Determination of the physiological dimer interface of the PhoQ sensor domain. *J Mol Biol* 379: 656-665.
25. Parkinson JS (2010) Signaling mechanisms of HAMP domains in chemoreceptors and sensor kinases. *Annu Rev Microbiol* 64: 101-122.
26. Batchelor E, Goulian M (2003) Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system. *Proceedings of the National Academy of Sciences of the United States of America* 100: 691-696.
27. Hulko M, Berndt F, Gruber M, Linder JU, Truffault V, et al. (2006) The HAMP domain structure implies helix rotation in transmembrane signaling. *Cell* 126: 929-940.
28. Marina A, Waldburger CD, Hendrickson WA (2005) Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *EMBO J* 24: 4247-4259.
29. MacKenzie KR, Prestegard JH, Engelman DM (1997) A transmembrane helix dimer: structure and implications. *Science* 276: 131-133.
30. Marassi FM, Opella SJ (2000) A solid-state NMR index of helical membrane protein structure and topology. *J Magn Reson* 144: 150-155.

31. Pakula AA, Simon MI (1992) Determination of transmembrane protein structure by disulfide cross-linking: the *Escherichia coli* Tar receptor. *Proc Natl Acad Sci U S A* 89: 4144-4148.
32. Lynch BA, Koshland DE, Jr. (1991) Disulfide cross-linking studies of the transmembrane regions of the aspartate sensory receptor of *Escherichia coli*. *Proc Natl Acad Sci U S A* 88: 10402-10406.
33. Soto CS, Hannigan BT, DeGrado WF (2011) A Photon-Free Approach to Transmembrane Protein Structure Determination. *Journal of Molecular Biology* 414: 596-610.
34. Goldberg SD, Clinthorne GD, Goulian M, DeGrado WF (2010) Transmembrane polar interactions are required for signaling in the *Escherichia coli* sensor kinase PhoQ. *Proc Natl Acad Sci U S A* 107: 8141-8146.
35. Moukhametzianov R, Klare JP, Efremov R, Baeken C, Goppner A, et al. (2006) Development of the signal in sensory rhodopsin and its transfer to the cognate transducer. *Nature* 440: 115-119.
36. Chang DK, Cheng SF, Trivedi VD, Lin KL (1999) Proline affects oligomerization of a coiled coil by inducing a kink in a long helix. *Journal of Structural Biology* 128: 270-279.
37. Yohannan S, Faham S, Yang D, Whitelegge JP, Bowie JU (2004) The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. *Proc Natl Acad Sci U S A* 101: 959-963.
38. Meruelo AD, Samish I, Bowie JU (2011) TMKink: a method to predict transmembrane helix kinks. *Protein Sci* 20: 1256-1264.
39. Senes A, Engel DE, DeGrado WF (2004) Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs. *Curr Opin Struct Biol* 14: 465-479.
40. Riek RP, Rigoutsos I, Novotny J, Graham RM (2001) Non-alpha-helical elements modulate polytopic membrane protein architecture. *J Mol Biol* 306: 349-362.
41. Rigoutsos I, Riek P, Graham RM, Novotny J (2003) Structural details (kinks and non- α conformations) in transmembrane helices are intrahelically determined and can be predicted by sequence pattern descriptors. *Nucleic acids research* 31: 4625.
42. Grigoryan G, Degrado WF (2011) Probing designability via a generalized model of helical bundle geometry. *J Mol Biol* 405: 1079-1100.
43. Crick FHC (1953) The Fourier Transform of a Coiled-Coil. *Acta Crystallographica* 6: 685-689.
44. Senes A, Chadi DC, Law PB, Walters RF, Nanda V, et al. (2007) E(z), a depth-dependent potential for assessing the energies of insertion of amino acid side-chains into membranes: derivation and applications to determining the orientation of transmembrane and interfacial helices. *J Mol Biol* 366: 436-448.
45. Airola MV, Watts KJ, Bilwes AM, Crane BR (2010) Structure of concatenated HAMP domains provides a mechanism for signal transduction. *Structure* 18: 436-448.

46. Amadei A, Linssen AB, Berendsen HJ (1993) Essential dynamics of proteins. *Proteins* 17: 412-425.
47. Pagni M, Ioannidis V, Cerutti L, Zahn-Zabal M, Jongeneel CV, et al. (2007) MyHits: improvements to an interactive resource for analyzing protein sequences. *Nucleic acids research* 35: W433-W437.
48. Pagni M, Ioannidis V, Cerutti L, Zahn-Zabal M, Jongeneel CV, et al. (2004) MyHits: a new interactive resource for protein annotation and domain identification. *Nucleic acids research* 32: W332-W335.
49. Zhou Q, Ames P, Parkinson JS (2009) Mutational analyses of HAMP helices suggest a dynamic bundle model of input-output signalling in chemoreceptors. *Molecular Microbiology* 73: 801-814.
50. Cheung J, Hendrickson WA (2009) Structural Analysis of Ligand Stimulation of the Histidine Kinase NarX. *Structure* 17: 190-201.
51. Laio A, Parrinello M (2002) Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America* 99: 12562-12566.
52. Falke JJ, Hazelbauer GL (2001) Transmembrane signaling in bacterial chemoreceptors. *Trends in Biochemical Sciences* 26: 257-265.
53. Hazelbauer GL (2001) Transmembrane signaling and adaptation in bacterial chemoreceptors. *Biophysical Journal* 80: 182A-182A.
54. Lee GF, Lebert MR, Lilly AA, Hazelbauer GL (1995) Transmembrane Signaling Characterized in Bacterial Chemoreceptors by Using Sulfhydryl Cross-Linking in-Vivo. *Proceedings of the National Academy of Sciences of the United States of America* 92: 3391-3395.
55. Regelman AG, Lesley JA, Mott C, Stokes L, Waldburger CD (2002) Mutational analysis of the *Escherichia coli* PhoQ sensor kinase: Differences with the *Salmonella enterica* serovar typhimurium PhoQ protein and in the mechanism of Mg²⁺ and Ca²⁺ sensing. *Journal of Bacteriology* 184: 5468-5478.
56. Carpenter T, Bond PJ, Khalid S, Sansom MS (2008) Self-assembly of a simple membrane protein: coarse-grained molecular dynamics simulations of the influenza M2 channel. *Biophysical Journal* 95: 3790-3801.
57. Prakash A, Janosi L, Doxastakis M (2010) Self-association of models of transmembrane domains of ErbB receptors in a lipid bilayer. *Biophysical Journal* 99: 3657-3665.
58. Periole X, Huber T, Marrink SJ, Sakmar TP (2007) G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. *Journal of the American Chemical Society* 129: 10126-10132.
59. Combet C, Blanchet C, Geourjon C, Deleage G (2000) NPS@: network protein sequence analysis. *Trends Biochem Sci* 25: 147-150.
60. Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* 36: W197-201.
61. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics* 14: 892-893.

62. Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C (2009) A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct Biol* 9: 51.
63. Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, et al. (2005) Protein structure prediction servers at University College London. *Nucleic Acids Res* 33: W36-38.
64. Rost B, Yachdav G, Liu J (2004) The PredictProtein server. *Nucleic Acids Res* 32: W321-326.
65. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14: 33-38, 27-38.
66. Henin J, Shinoda W, Klein ML (2008) United-atom acyl chains for CHARMM phospholipids. *J Phys Chem B* 112: 7008-7015.
67. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 26: 1781-1802.
68. Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr., Nilsson L, Petrella RJ, et al. (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30: 1545-1614.
69. Barth P, Schonbrun J, Baker D (2007) Toward high-resolution prediction and design of transmembrane helical protein structures. *Proc Natl Acad Sci U S A* 104: 15682-15687.
70. Bass RB, Butler SL, Chervitz SA, Gloor SL, Falke JJ (2007) Use of site-directed cysteine and disulfide chemistry to probe protein structure and dynamics: applications to soluble and transmembrane receptors of bacterial chemotaxis. *Methods Enzymol* 423: 25-51.
71. Clinthorne GD (2012) PhoQ: Structural and Mechanistic investigations into an important bacterial sensor kinase. Philadelphia: University of Pennsylvania
72. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 79: 926.
73. Brunger A, Brooks CL (1984) Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chemical Physics Letters* 105: 495-500.
74. Feller SE, Zhang Y, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation: the Langevin piston method. *The Journal of Chemical Physics* 103: 4613.
75. Viklund H, Bernsel A, Skwark M, Elofsson A (2008) SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. *Bioinformatics* 24: 2928-2929.
76. Viklund H, Elofsson A (2008) OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics* 24: 1662-1668.

77. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305: 567-580.
78. Tusnady GE, Simon I (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* 17: 849-850.
79. Hofmann K (1993) TMbase-A database of membrane spanning protein segments. *Biol Chem Hoppe-Seyler* 374: 166.
80. Hennerdal A, Elofsson A (2011) Rapid membrane protein topology prediction. *Bioinformatics* 27: 1322-1323.
81. Cserzo M, Wallin E, Simon I, von Heijne G, Elofsson A (1997) Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein engineering* 10: 673-676.

3.3 Near-atomistic model of the complete PhoQ histidine kinase

The cytoplasmic domain of PhoQ is largely unsolved. MODELLER, a comparative modeling software for three-dimensional protein structures, was used to assemble the complete atomistic model of the PhoQ homodimer [1]. The crystal structure of the sensor (PDB: 3bq8, [2]) and catalytic (PDB: 1id0, [3]) domain have been solved for *E. coli*. The histidine kinase (PDB: 2C2A, identity: 22.9%, [4]) and HAMP (PDB: 2I7h, identity: 31.8%, [5]) domains of, respectively, *Thermotoga maritima* and *Archaeoglobus fulgidus* were used as templates for the cytoplasmic domain. The preserved residues are mostly located in the helical domains, supporting a probable functional importance. The TM domain was taken from cross-linking guided models produced by using MD simulations (see Section 3.2).

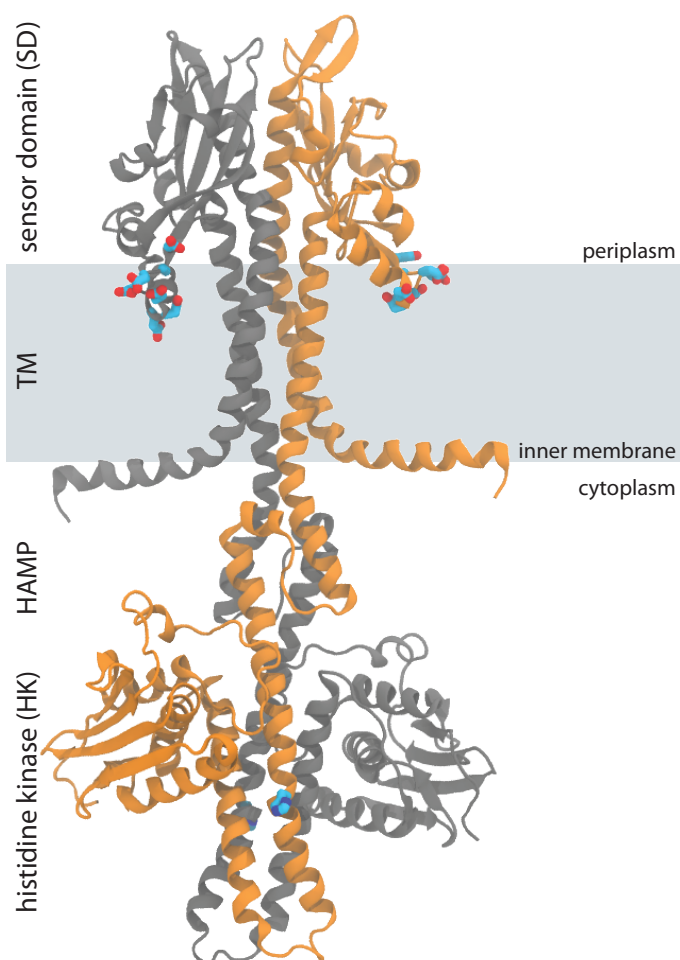


Figure 10: Atomistic model of PhoQ. The acid cluster and H277 are highlighted in licorice representation.

In the final model, the large acidic cluster (148—EDDDDAE—154) of the sensor domain (SD), involved in sensing divalent cations and AMP, lies very close to the

membrane surface (Figure 10). It was proposed that the SD senses the electrostatic variation at the bacterial membrane surface [6]. When the concentration of Mg^{2+} or Ca^{2+} cations is high (> 1 mM), cations can bind to the acidic cluster, forming a metal-mediated bridge with the negatively-charged membrane of bacteria. Based on this assumption and consistent with the proposed model, each SD acidic cluster could closely interact with the membrane surface [6]. The repulsion or attraction of the acidic cluster may lead to a scissoring movement at the SD interface. Such a movement could induce a conformational change of the transmembrane domain, consistent with the one observed in MD. Finally, the resulting rearrangement of the HAMP and histidine kinase domains would trigger the phosphorylation of the His277 and a phosphotransfer to PhoP.

3.3.1 References

1. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, et al. (2006) Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* Chapter 5: Unit 5 6.
2. Cheung J, Bingman CA, Reyngold M, Hendrickson WA, Waldburger CD (2008) Crystal structure of a functional dimer of the PhoQ sensor domain. *J Biol Chem* 283: 13762-13770.
3. Marina A, Mott C, Auyzenberg A, Hendrickson WA, Waldburger CD (2001) Structural and mutational analysis of the PhoQ histidine kinase catalytic domain. Insight into the reaction mechanism. *J Biol Chem* 276: 41182-41190.
4. Marina A, Waldburger CD, Hendrickson WA (2005) Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *EMBO J* 24: 4247-4259.
5. Ferris HU, Dunin-Horkawicz S, Mondejar LG, Hulko M, Hantke K, et al. (2011) The mechanisms of HAMP-mediated signaling in transmembrane receptors. *Structure* 19: 378-385.
6. Cho US, Bader MW, Amaya MF, Daley ME, Klevit RE, et al. (2006) Metal bridges between the PhoQ sensor domain and the membrane regulate transmembrane signaling. *J Mol Biol* 356: 1193-1206.

3.4 Modeling of cardiolipin species

Published as the following paper:

“Cardiolipin models for molecular simulations of bacterial and mitochondrial membranes”,

Lemmin T., Bovigny C., Lançon D., and Dal Peraro M., Journal of Chemical Theory and Computation, 2012, 9 (1), pp 670–678

3.4.1 INTRODUCTION

The biological membrane constitutes an essential barrier for all living organisms. It is composed of a complex mixture of lipids and embedded proteins, forming a heterogeneous liquid-crystalline structure, organized into rafts of local association of lipids, with specific physicochemical properties [1]. Over the past years, it has been established that lipids do not constitute a simple structural support for proteins, but interact, modify and regulate their function [2,3]. Cells invest a substantial amount of energy to synthesize a large repertoire of different specific lipids, and nearly 5% of genes are involved in this synthesis [4].

The composition of cell membranes in archaea, eukaryotes and bacteria differ significantly. In particular, cardiolipins (CLs) are anionic phospholipids found in Gram-negative and Gram-positive bacteria, as well as in mitochondria and chloroplast inner membranes of eukaryotes. CLs are composed of two phosphatidylglycerols moieties connected by a single glycerol molecule (Figure 11). This dimeric structure is unique among phospholipids, and confers an important role in the structural stabilization and activation membrane-dependent processes [5,6]. CLs can potentially carry two negative charges, one on each phosphate group (Figure 11c). Despite the apparent symmetry of the head group, the ionization of the phosphate groups occurs at different acidity levels ($pK_1 \sim 2.8$ and $pK_2 \sim 7.5 - 9.5$) [7]. Under physiological conditions, CLs may therefore carry one negative charge. However, depending on the local pH and CL concentration, the deprotonation of both head groups cannot be excluded [8]. Importantly, cardiolipins have the tendency to produce high membrane curvature due to their large hydrophobic region and small charged head. Moreover, their bilayer and non-bilayer lipid properties are strongly modulated by the presence of divalent cations, namely Mg^{2+} and Ca^{2+} ions.

Despite the importance of CLs in bacterial and mitochondrial membrane composition and their key role in cell division [9], energy metabolism and pathologies [10], there are to date few molecular, biophysical, and biochemical characterizations thoroughly describing their properties and specific interactions with other lipid species and protein complexes.

Molecular simulations, at different levels of resolution, have contributed in many respects to describe the structural and kinetic properties of biological membranes in the last two decades [11,12,13]. Coarse-grained and united-atom models for cardiolipins have been parameterized [14,15,16,17]. These models provided useful mechanistic insights into phase transitions and aggregate geometries, but the lack of atomistic details inherent to a coarse-grained approach prevents a full molecular understanding of their properties. Similarly, united-atom models may suffer from the

same limitations [14]. Rog *et al.* proposed [18] the first all-atom OPLS-based model for the study of the mitochondrial inner membrane. Only the doubly charged protonation state was modeled in this case and the partial charges were directly derived from Murzyn *et al.* and adapted from the OPLS force field [19]. They observed that the effect of CLs on the membrane properties is strongly dependent on the membrane composition and cannot be easily predicted. The interaction with cations, which is known to influence CL properties, however, has not yet been described. Recently, Aguayo *et al.* [16] proposed a new hybrid set of parameters for the doubly charged cardiolipins in the united-atom CHARMM27-UA and all-atom CHARMM36 framework. They reported the effect of cardiolipins on the order of the acyl chains and membrane thickness, which can be associated to mitochondrial membrane activity.

To date, atomistic models of all the possible protonation states of CLs have not yet been parameterized for classical molecular dynamics (MD) simulations consistent with the commonly used AMBER or CHARMM force fields. Therefore, we propose in this work an *ab initio* parameterization of CL models with different physiological protonation states, in which the phosphate groups can either be unprotonated or singly protonated. In the singly protonated CL, a proton is trapped in the bicyclic resonance structure formed by the two phosphate and hydroxyl groups (Figure 11c), for which an energy barrier as low as 4 – 5 kcal/mol has been estimated for the proton exchange between the phosphates [20].

We used and tested these initial models to characterize CL-rich membrane patches and the interaction of CLs with monovalent and divalent cations. Our results were consistently validated *a posteriori* by quantum mechanics calculations and available experimental data, such as X-ray structures of proteins co-crystallized with cardiolipin molecules.

3.4.2 RESULTS

CL structure and parameterization were obtained following the procedure given in the Material and Methods section. This set of parameters is the result of an extensive conformational search of CL head group space, from which the most representative structures were clustered and optimized to produce a novel and consistent set of partial atomic charges. Since the pK_2 of cardiolipins (>7.5) is of the same order as the physiological pH, this would imply the coexistence of different protonation states of the phosphate groups. Thus, two protonation states for cardiolipin head groups were considered in this study: CLs with unprotonated phosphate groups (uCL) and a singly protonated phosphate group (pCL) (Figure 11c). To characterize the effect of the protonation, three cardiolipin bilayers were assembled. The first two systems were exclusively composed of uCL or pCL molecules. The third system was a uniform mixture of both kinds of cardiolipins. Molecular dynamics simulations were carried out for all three systems using CL parameterization consistent with the CHARMM and AMBER force fields. The subscripts “a” and “c” will be used to differentiate simulations carried out with the AMBER and CHARMM force fields, respectively (e.g. uCL_a: unprotonated CL with AMBER force field). CL refers to the cardiolipin independent of its protonation state.

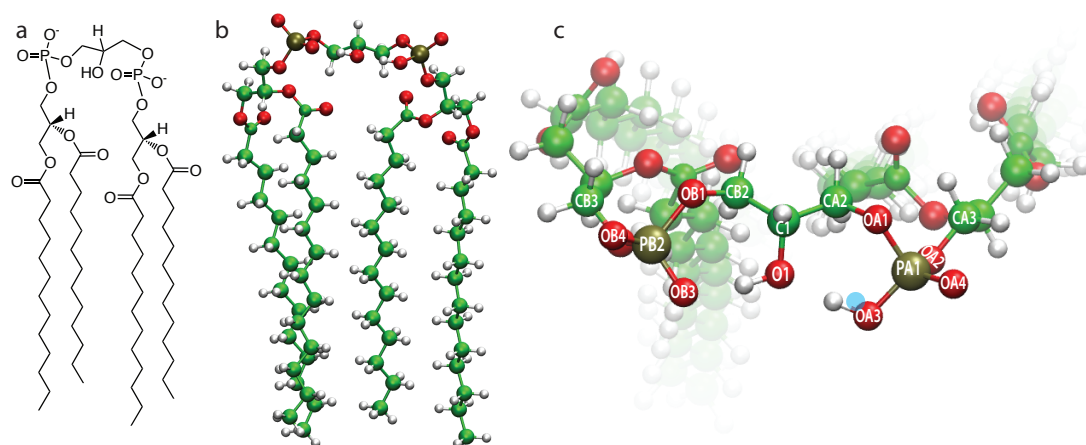


Figure 11: Cardiolipin structure and protonation states. Cardiolipin (CL) chemical structure (a) and 3D model (b). The cardiolipin forms a unique dimeric phospholipid composed of four acyl tails. (c) CLs can be unprotonated (uCL) or singly protonated (pCL) depending on the protonation state of the OA3 oxygen atom, highlighted in blue. Relevant atom names composing CL head group are shown on the 3D structure, consistently with the topology and parameter files reported in Supporting Informations.

Biophysical characterization of the cardiolipin bilayer

Experimental data showed that the cardiolipin phase transition temperature is significantly higher than those corresponding to similar diacyl phospholipids [21]. Consistent with these results, a membrane patch composed exclusively of cardiolipins arranges in a gel phase conformation. To test our models, we quantified the order parameter (S_{CD}), which represents a measure of the orientation mobility of carbon-deuterium bonds that can be observed in the 2H NMR experiments (Figure 12a). The computed S_{CD} profiles are in good agreement with the experimental data

for the hydrophobic region of the phospholipid membrane [22]. The protonation of the phosphate group slightly affects S_{CD} ; in pCL systems the order parameter is systematically higher. The higher order parameter for CL_a indicates that the acyl chains are more parallel to the bilayer normal when using the AMBER parameterization. Mixing cardiolipins with different protonation states only marginally affected the order parameter, which remains comparable to the uniform CL patches.

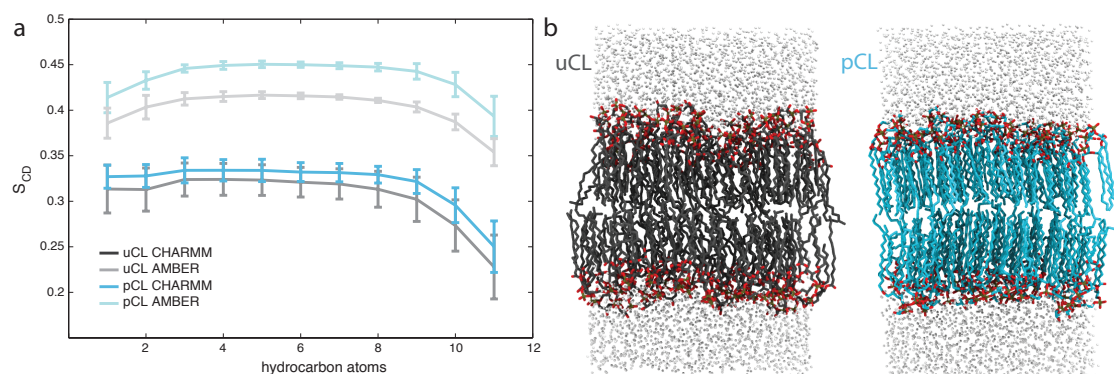


Figure 12: Biophysical properties of CL-rich membrane patches. (a) SCD order parameters for pCL and uCL for both the AMBER and CHARMM parameterization. (b) Side view MD snapshots of the CL membrane systems for pCL and uCL simulated using the CHARMM force field.

Another consequence of the difference between the two parameterizations is seen by the thickness of the membranes. Membranes simulated with the AMBER force field are thicker compared to the CHARMM simulations, ~ 45 Å and ~ 42 Å, respectively (Table 1). At 35° C, the thickness measured by x-ray diffraction was estimated to be ~ 43 Å [23]. The acyl tails are more extended and vertical in the simulation performed with the AMBER force field and explains this difference. It was observed that MD simulations of a membrane using AMBER parameters have a tendency to favor gel over liquid crystal phase. This may be due to the still limited testing of the force field parameters for lipids [11,12].

Table 1. Biophysical properties of cardiolipin membrane patches simulated in this study.

System	Force field	Thickness [Å]	Area per lipid [†] [Å ²]		Volume per lipid [Å ³]		Diffusion coefficient [10 ⁻⁸ cm ² /s]
CDL Na ⁺	CHARMM	41.46 ± 1.52	90.81	± 0.20	1825.43 ± 12.68	4.85 ± 0.15	
			(30.25)				
CDL Mg ²⁺	CHARMM	39.53 ± 2.81	100.03	± 0.21	1789.59 ± 12.41	4.65 ± 0.25	
			(27.41)				
	AMBER	43.26 ± 2.6	86.45	± 0.18	1859.77 ± 15.78	2.46 ± 0.13	
			(16.16)				
CLP Na ⁺	CHARMM	42.67 ± 2.29	86.97	± 0.17	1808.66 ± 11.54	5.26 ± 0.46	
			(25.46)				
CLP Mg ²⁺	CHARMM	40.93 ± 2.63	92.83	± 0.19	1795.58 ± 12.60	5.11 ± 0.18	
			(28.16)				
	AMBER	44.18 ± 2.95	82.93	± 0.14	1872.08 ± 16.97	2.67 ± 0.24	
			(21.48)				
mixed Na ⁺	CHARMM	41.60 ± 2.56	90.28	± 0.18	1818.10 ± 13.12	5.02 ± 0.08	
			(24.40)				
mixed Mg ²⁺	CHARMM	41.42 ± 2.42	92.74	± 0.20	1784.01 ± 12.70	4.54 ± 0.45	
			(35.50)				
	AMBER	44.74 ± 2.96	82.54	± 0.16	1833.82 ± 17.62	2.12 ± 0.11	
			(20.61)				

[†]The value in parenthesis indicates the standard deviation measured for the patch.

In order to estimate the variation within the membrane patch, the area per lipid was estimated using a Voronoi decomposition applied to the position of the pairs of CL phosphorus atoms [24]. The order parameters would correspond to bilayer in the gel phase, whereas the CL area per lipid is close to the one measured in liquid crystal phase (83 \AA^2 at 35°C) [23]. A similar difference has been observed for the recently reported united-atom parameterization of cardiolipins [16]. The pCLs is significantly packed tighter than uCLs. The lipids are more tightly packed for the simulations carried out with the AMBER force field independent of the protonation state. These values are in good agreement with the ones reported by Aguayo *et al.* [16], which are the closest models reported in the literature to our parameterization. On the other hand, when mixing different protonation states, the area per lipid is comparable to the pure uCL patch. The variation of the area per lipid within a patch remained comparable for all the systems.

The diffusion coefficient of CL_c is in good agreement with the expected values for lipids from the experimental data ($\sim 4 \cdot 10^{-8} \text{ cm}^2/\text{s}$) [25]. The pCL_c diffuse faster than the unprotonated cardiolipins (Table 1). The CL_a diffuses slower ($1.38 \pm 0.01 \cdot 10^{-8} \text{ cm}^2/\text{s}$), which is consistent with the more significant gel-like conformation adopted by the acyl tails.

Exploring the conformational space of the cardiolipin polar heads

Experimental data point to a predominant functional role of cardiolipins in several crucial cellular mechanisms, such as protein regulation, energy metabolism and structural membrane organization. Accurately reproducing the conformation space of the polar heads is therefore essential. Currently, only limited information is available for CL head group geometries.

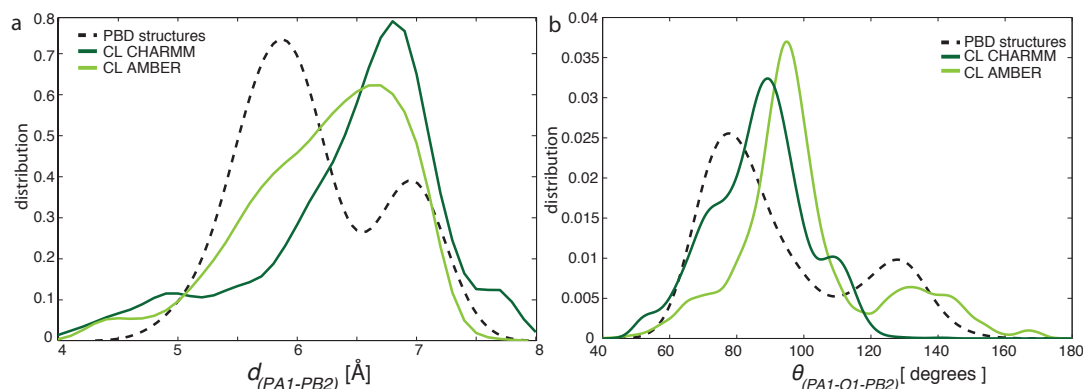


Figure 13: Conformational flexibility of the CL head groups. (a) The inter-phosphate distance $d(\text{PA1-PB2})$ and (b) phosphate-hydroxyl angle $\theta(\text{PA1-O1-PB2})$ of the polar heads are shown. Since the protonation state did not affect these variables, the simulations are averaged depending on the force field used. The relative quantities extracted from CL crystallized in the PDB structures are shown with dashed lines (see SI for the exhaustive list of the protein complexes used for this analysis).

We therefore decided to characterize these properties based on our models and compare and validate the various conformers extracted from the MD simulations with the crystallographic structures of CL from PDB data, as proposed by Dahlberg *et al.* [14]. These structures were co-crystallized with proteins and are therefore not *per se* fully representative of a membrane bilayer. We expected however that they would sample the head group conformational space to some extent. A total of 91 structures formed the set of crystal structures (see Table S1 in the Supporting Material section). The head group geometry was characterized based on the inter-phosphate distance $d_{(\text{PA1}-\text{PB2})}$ and the phosphate-hydroxyl angle $\theta_{(\text{PA1}-\text{O1}-\text{PB2})}$ (Figure 13). These two variables reduce the degrees of freedom of the conformational space and describe global characteristics of the head group geometry.

The PDB inter-phosphate distance is characterized by a bimodal distribution, with a first maximum at 5.9 Å and the other at 7.0 Å (Figure 13a). In the molecular dynamics simulations, the maximum is situated at 6.8 Å and 6.7 Å for AMBER and CHARMM, respectively. The maximum for both force fields is closer to the second experimental maximum (7.0 Å). CHARMM covers more uniformly the conformational space. Despite this shift, the conformation space explored by both models is in good agreement with these experimental data, even though the PDB structures do not represent bilayer structures. The protonation state does not significantly influence these distributions. The angle distribution is also bimodal (maxima at 78 and 128 degrees). CHARMM-based models show a single mode distribution, which largely overlaps the PDB global maxima. The principal mode for the AMBER force field is shifted (95 degrees), but the distribution is bimodal with a second mode closer to PDB (132 degrees; Figure 13b).

In order to further quantify the accuracy of the atomic partial charges and dihedral parameters derived for the head group, a hybrid quantum mechanics/molecular mechanics (QM/MM) geometry optimization was carried out. A set of 50 membrane bilayer structures was extracted from the MD simulations. Single CL molecules in the patch were randomly selected to be geometrically optimized at the QM/MM level. The refined structures were in good agreement with the MD head group conformation (RMSD 0.13 ± 0.03 Å). The inter-phosphate distance and phosphate-hydroxyl angle varied only slightly (0.08 ± 0.13 Å and 0.39 ± 3.00 degrees, respectively). The angles and dihedrals of the bonded atoms of the head group remained constant between the MM and QM structures with a variation under 10%.

Therefore, the reasonably good conformational overlap with available crystallized structures of CLs, and the limited variation of CL geometries relaxed at a higher level of theory support the structural robustness of the force field parameters derived here for MD simulations.

Cardiolipin interactions with cations

Cations are known to be crucial for the structural and functional integrity of cell membranes. The anionic nature of cardiolipins makes them a major binding site for cations, or positively charged species (*e.g.* AMPs). We therefore analyzed the distribution along the z-axis normal to the membrane surface and the radial distribution function of Na^+ and Mg^{2+} with respect to CL lipids (Figure 14 and Figure 15). A larger patch (200×200 Å²) was also used to study the combined effect of Na^+

and Mg^{2+} . Since at this stage, the CHARMM model seems to better capture the physical properties of the membrane, the results for this model will be presented in more detail (additional results for AMBER models are reported in the Supporting Information section).

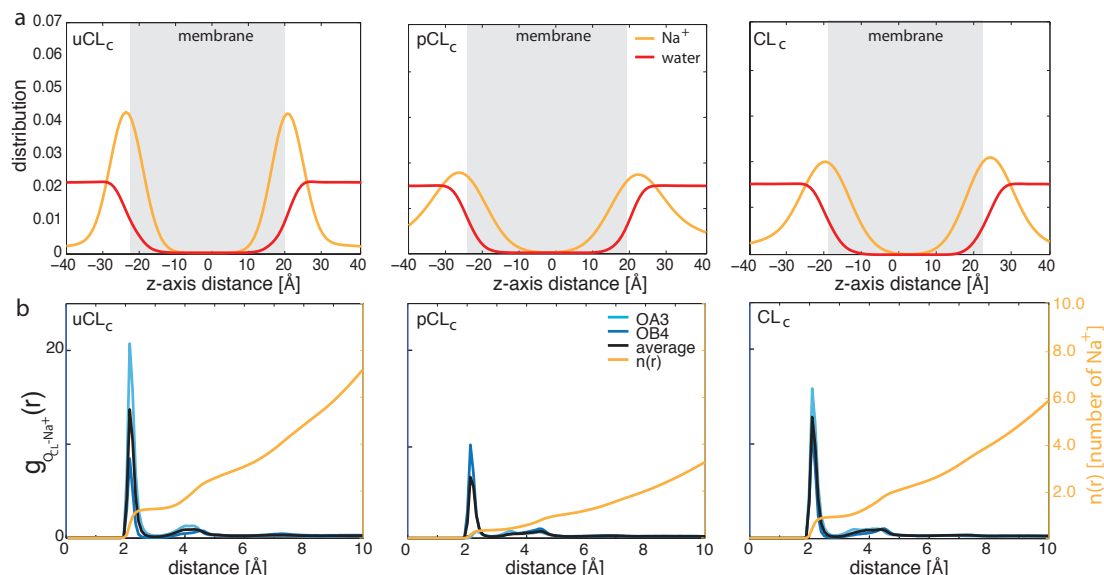


Figure 14: Cardiolipin interactions with sodium. (a) Distribution of Na^+ (orange) and water molecules (red) along the z-axis normal to the membrane surface. From left to right, data for the uCL, pCL and CL (mixed uCL/pCL) systems are reported for the CHARMM force field. (b) Radial distribution function $g(r)$ and coordination number $n(r)$ (orange) for Na^+ with respect to CL phosphate oxygen atoms for simulations carried out with CHARMM force field.

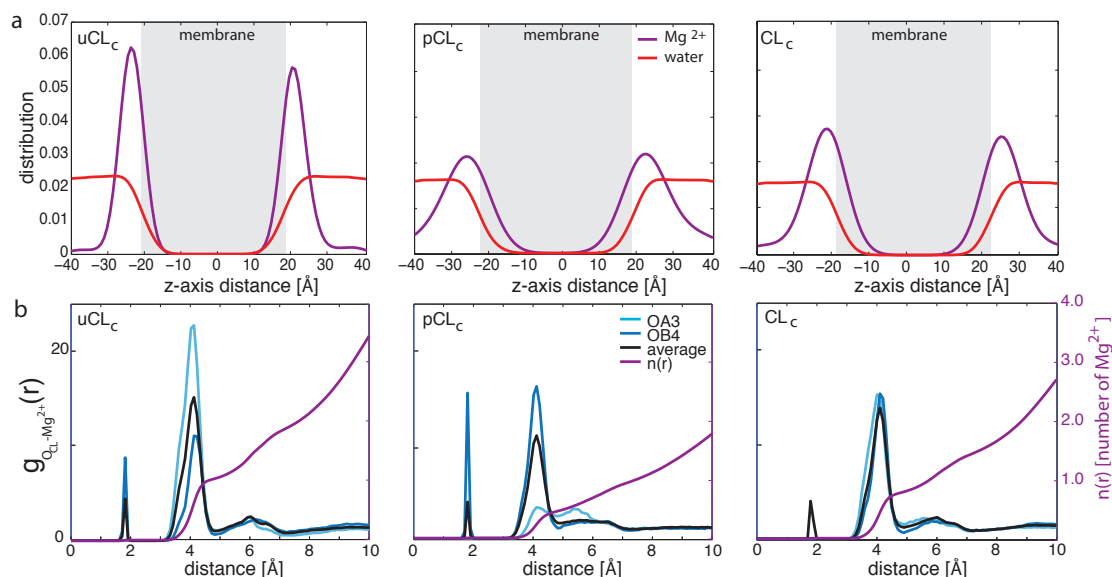


Figure 15: Cardiolipin interactions with magnesium. (a) Distribution of Mg^{2+} (purple) and water molecules (red) along the z-axis normal to the membrane surface. From left to right, data for the uCL, pCL and CL (mixed uCL/pCL) systems are reported for the CHARMM force field. (b) Radial distribution function $g(r)$ and coordination number $n(r)$ (purple) for Mg^{2+} with

respect to phosphate oxygen atoms for simulations carried out with CHARMM force field (see results for AMBER force field in SI section).

In the unprotonated system, 90% of the Na^+ and 85% of the Mg^{2+} are localized within 4 Å from the membrane surface (Figure 14a and Figure 15a, respectively). The protonation leads to a broadening of the distribution, reducing to 55% the number of Na^+ and 76% the number of Mg^{2+} within 4 Å from the surface. It is of interest to note that this proportion increases to ~ 89% for both cations when interacting with the mixed protonation state bilayer. Consistent with the experimental data, the phosphate groups constitute the main binding site for cations. On one hand, Na^+ cations are partially buried in to the membrane (Figure 14a). These cations interact directly with the CL phosphate oxygen atoms (i.e. OA3 and OB4), corresponding to the first peak in the radial distribution function (Figure 14b). Several Na^+ cations interact with a single cardiolipin molecule (Figure S6, for a representative CL – Na^+ conformation). On the other hand, despite a few Mg^{2+} cations form direct interactions with the CL phosphate oxygen atoms, the main solvation structure indicates that CL – divalent cation interactions are water mediated. The most relevant peak found at ~ 4 Å corresponds to Mg^{2+} surrounded by their first solvation shell in an octahedral conformation, with water hydrogen atoms pointing towards the phosphate group or the bulk. The peak at ~ 6 Å mainly represents the other phosphate group of a cardiolipin, which interacts with a cation (see Figure S6 for a representative CL – Mg^{2+} conformation). For the unprotonated systems (uCL), each phosphate is able to interact globally with one cation in the first shell, whereas when one phosphate is protonated (pCL systems), the cation occupancy is nearly half (pCL system in Figure 15 and Figure S5b). When the uCL and pCL moieties are mixed, the phosphate-cation interactions are mixed as well, showing an improved ability, as compared to the pCL system, to globally attract cations close to the membrane surface. The area per lipid decreases significantly in presence of Na^+ in the pure protonated or unprotonated cardiolipin patches (Table 1). The order parameters, however, are not modified and therefore do not contribute to this difference in the area per lipid. The effect of cations on the area per lipid is only marginal in the mixed protonated CL bilayer.

The different CL protonation states produce a different electrostatic signature for the membrane surface, which is directly related to the CL ability to interact with cations and possibly other proteins and/or phospholipids. As expected, the uCL membrane surface displays a distinct strong negative charge, whereas the surface for pCL is less charged. This is consistent with the normal distribution of cations at the membrane surface (Figure 16). Cations are localized in contact with the most highly charged regions. Complex and intricate networks involving several cardiolipins were observed. In presence of monovalent cations, the cardiolipins form a very tight network trapping up to three Na^+ per CL (Figure S6). In contrast, the cardiolipins form pockets composed of up to three lipids that are able to trap a Mg^{2+} cation (Figure S6).

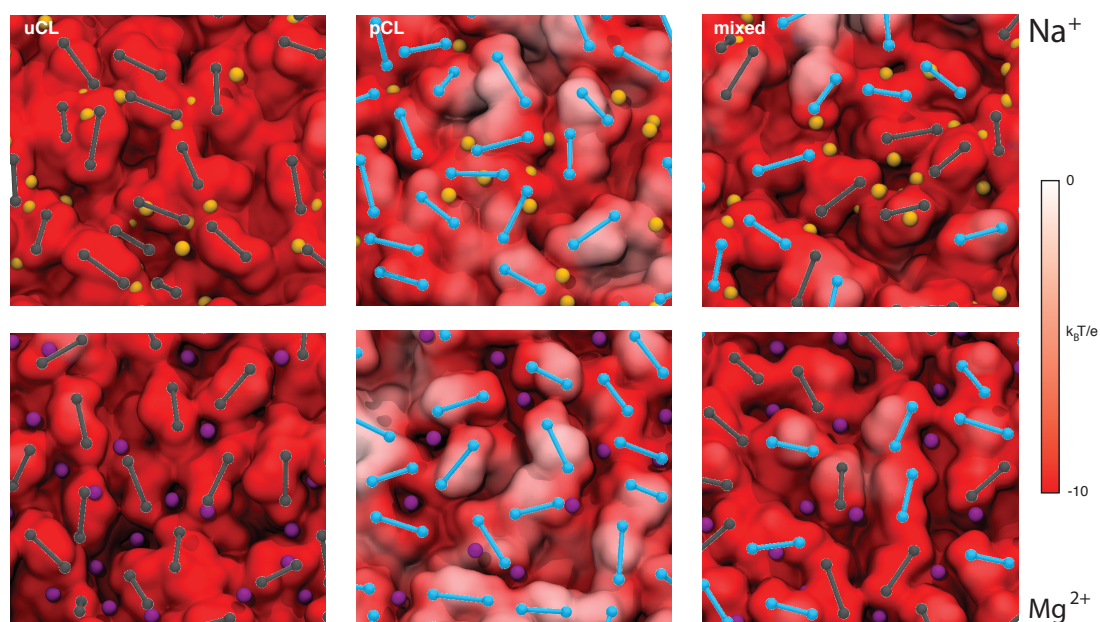


Figure 16: Electrostatic properties of the CL-rich membrane surface. The electrostatic potential of the cardiolipin membrane surface was computed with APBS [26] and plotted on the molecular surface of the CL membrane. For sake of clarity, the uCL molecules are topologically identified with gray balls representing the phosphorus atoms connected by sticks. Similarly, pCL moieties are in blue. Na^+ and Mg^{2+} cations are shown as orange and purple spheres, respectively.

When Na^+ and Mg^{2+} cations are both present in a mixed protonated CL bilayer, more cations are simultaneously interacting with at least one protonated and one unprotonated cardiolipin. There is up to twice as many Na^+ than Mg^{2+} interacting only with uCL molecules and at the interface of the protonated and unprotonated cardiolipins (Figure 17a). In contrast, the pCL interacts almost equally with both cations. The radial distribution of Mg^{2+} within 6 Å of the membrane surface corresponds to the one expected for molecules which do not interact (Figure 17b). This indicates a random distribution of Mg^{2+} on the membrane surface. In contrast, Mg^{2+} radial distribution function with respect to Na^+ shows a clear peak at 5 Å, which corresponds to Mg^{2+} and Na^+ cations interacting simultaneously with the same cardiolipin molecules. The cardiolipins impose a structural order to the Na^+ cations, via its direct interaction with the phosphate group. The interaction with the Na^+ packs the CL head groups tighter, allowing a Mg^{2+} cation to be trapped above the Na^+ (Figure S6). On the average, such an interaction involves a single Mg^{2+} and Na^+ cation and represents $\sim 5\%$ of Mg^{2+} localized close to the membrane surface.

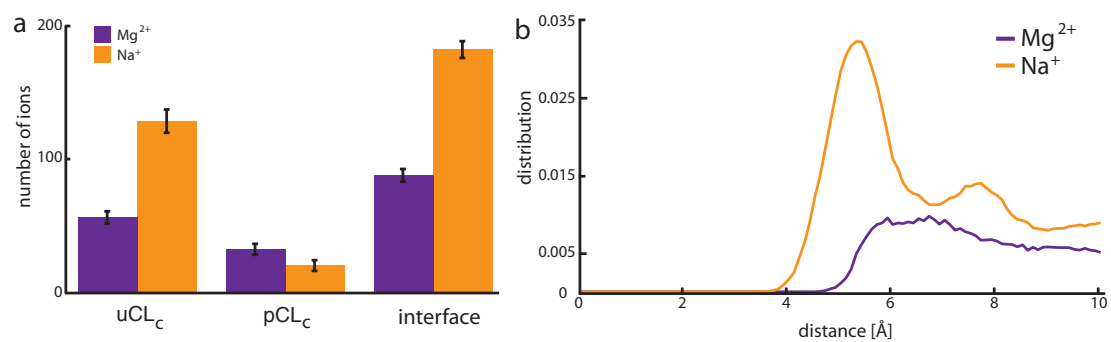


Figure 17: Cardiolipins in the presence of sodium and magnesium cations. (a) Number of cations interacting with only uCL molecules, pCL molecules and with at least one uCL and one pCL molecule (interface), and (b) Radial distribution function $g(r)$ for Mg^{2+} with respect to Mg^{2+} and Na^+ .

3.4.3 DISCUSSION

Cardiolipins represent a unique anionic lipid species, which is only present in bacterial and mitochondrial membranes, and plays a key role in several cellular functions. To date only limited information concerning CL physicochemical properties is available. We therefore built and parameterized cardiolipin models for the commonly used AMBER and CHARMM force fields, to allow a further investigation of their fundamental role through molecular dynamics simulations. Since the pK_2 of CLs is close to the physiological pH, two protonation states for the phosphate groups are likely to be present under normal conditions (an unprotonated (uCL) and the singly protonated (pCL) CL conformation); their properties, mutual interactions and interactions with cations are characterized for the first time in this study.

Our model accurately reproduces the lateral diffusion coefficient and the area per lipid expected for lipid bilayers, with the CHARMM parameterization better capturing the physical properties of the membrane. This difference can be explained by the fact that AMBER acyl parameters have not yet been extensively tested and optimized for lipid bilayers. Recently, a new set of parameters for the acyl chains has been proposed [27]. Coupling these new parameters to our AMBER models of the head group can be a good starting point for further development.

The asymmetry created by the protonation of the phosphate groups leads to a modification of the area per lipid (Table 1). Consistent with our observations, Khalifat *et al.* studied the effect of pH on the packing of cardiolipin-containing bilayer [28], and proposed that the change in area per lipid is the driving force for membrane shape instability. Moreover, the presence of cardiolipin domains and their pH-dependent packing could function as a proton sink. It has been suggested in fact that cardiolipins stabilize the proton gradient across the membrane necessary for ATP synthesis [29]. Since cardiolipins are also associated to the proton uptake pathway in the cytochrome bc1 complex, they may play a role in proton conduction [30]. The proposed models will therefore allow investigating the effect of CL protonation on these systems at the molecular mechanics level.

A PDB data-mining was used to characterize the conformation space of cardiolipins. Event though, cardiolipins present in the PDB co-crystallized with proteins have a low statistical weight, and are not entirely representative of a bilayer conformation, the inter-phosphate distance and hydroxyl-phosphate angle measured from MD simulations nicely overlap with this distribution. This implies that the conformational space accessible to CL molecules forming a bilayer is greater, but largely overlaps with the one representing the interaction of CL with protein complexes. CLs thus can smoothly transition from the membrane phase to interacting with proteins without a specific contrasting conformational rearrangement. The higher CL binding affinity for specific protein regions would therefore be the only main driving force for CL-protein interactions.

It is known that cations, especially divalent, influence the physicochemical properties of cardiolipins (e.g. transition temperature, non-lamellar macroscopic phases) [21]. The CL surface displays a net negative charge attracting the cations to the membrane surface. Consistent with experimental evidence, we observe that cations

interact primarily with CL phosphate groups [31,32,33]. Although the monovalent cations interact directly with the oxygen atoms, only few divalent cations are trapped in the first solvation shell of the phosphate. Their main mode of interaction is through water-mediated bridges with CL phosphates. Nonetheless, CL-rich membranes are able to attract a high concentration of cations from the bulk. The cations neutralize up to 96% of the membrane surface charge. Thus, the binding of the cations could modify the membrane surface electrostatic potential. Using NMR, Macdonald *et al.* [34] showed that the binding of Ca^{2+} is predominantly due to electrostatic effects, rather than to the intrinsic affinity with CLs. Furthermore, in addition to the screening effect, interactions with cations shape the complex network of cardiolipins, which form negatively charged cavities where the Na^+ and Mg^{2+} cations are trapped.

In a bacteria membrane, phosphatidylglycerol (PG) and CL are the major contributors to the membrane surface negative charge. Antimicrobial peptides (AMPs) are part of the immune system, and are characterized by a cationic amphiphilic nature, mirroring the properties of the bacterial membrane. During critical phases of an infection, such as cell division, the bacterial membranes are enriched in cardiolipins at the expense of PG [35]. Cardiolipins with their mixed protonation conformations could therefore constitute a key target for the AMPs. It has been shown in fact that cardiolipins and divalent cations modulate the binding of antimicrobial peptides to Gram-positive bacteria [36].

In conclusion, the proposed models of cardiolipins have shown to be reliable and robust, and can reproduce the biophysical properties of cardiolipin-rich membrane portions. The availability of parameterized models accounting for multiple protonation states under physiological conditions will permit to build more realistic systems representative of the bacterial and mitochondrial membrane. This will allow gaining insight, through molecular simulations, into the biophysical properties of different membrane mixtures exposed to different environmental conditions, such as cation concentrations, AMP stress, and protein complexes interactions.

3.4.4 MATERIALS AND METHODS

Structural models and parameterization of the cardiolipin polar heads

The cardiolipin head group possesses two phosphate groups and can therefore carry up to two negative charges. CL pK_2 (>7.5) is of the order of the physiological pH, therefore two protonation states for the cardiolipin head group were considered for building the initial models: CLs with unprotonated phosphate groups (uCL) and a singly protonated phosphate group (pCL) (Figure 11). Ideal head groups were initially built considering atoms between the phosphate groups (from atoms OA2 to OB2 in Figure 11c). Two methyl molecules were used to cap the head group, replacing atom CA3 and CB3 as shown in Figure 11c.

An extensive search of the CL head group conformational space was carried out using Maestro[37]. A set of 991 conformers was generated. After minimization, 298 structures were clustered and ranked according to the OPLS force field [38] based energy function. The twenty most representative structures were finally extracted and an *ab initio* geometry optimization was performed with Gaussian09 [39] at the B3LYP/6-311+G**[40] level of theory. For the CHARMM-consistent parameterization, the partial atomic charges were computed at RHF/6-31G* level of theory and optimized to reproduce the water interaction using the Force Field Toolkit plug-in[41]. For the AMBER models, the partial atomic charges were instead extracted and averaged from single point calculations at the HF/6-31G* level of theory.

Finally, the atom types and bonded force constants were extracted directly from CHARMM36 [42] and general AMBER force field (GAFF) [43] parameter sets, respectively. This strategy ensures an optimal compatibility with both of the existing parameterizations. We completed the CL head models using the existing CHARMM36 parameterization for the acyl chains. The AMBER acyl parameters were instead extracted from the GLYCAM06 set [44]. These parameters, however, have not yet been finely tuned to reproduce the characteristics of acyl chains of phospholipids. The formatted topology and parameter files for MD simulations with the AMBER and CHARMM force fields are available in the Supporting Material.

Molecular dynamics simulations

We used molecular dynamics (MD) simulations to characterize the structural and dynamical properties of cardiolipin bilayers. In particular, tetramyristoyl cardiolipins were used for this study (Figure 11). Cardiolipins were assembled into a $60 \times 60 \text{ \AA}^2$ membrane patch composed of 50 lipids. The systems were then solvated in a $60 \times 60 \times 80 \text{ \AA}^3$ water box.

Na^+ and Mg^{2+} were used to investigate the effect of monovalent and divalent cations on CL. Mg^{2+} is an abundant cation in nature and plays a central role in several biological processes. For example, Mg^{2+} is essential for the maintenance of the bacterial membrane and is required for the synthesis of cardiolipins [45]. Cardiolipins are also responsible for the Mg^{2+} inhibition of SecA [46]. NaCl or MgCl_2 salts were added at a concentration of $\sim 150 \text{ mM}$. A larger $200 \times 200 \text{ \AA}^2$ cardiolipin patch composed of 450 lipids was assembled to study the ions interactions with CL when NaCl or MgCl_2 are mixed at a concentration of $\sim 150 \text{ mM}$.

All simulations were performed using NAMD [47] engine, with the CHARMM36 or AMBER ff99SB force fields [48]. TIP3P water [49] parameterization was used to describe the water molecules. The periodic electrostatic interactions were computed using the particle-mesh Ewald (PME) summation with a grid spacing smaller than 1 Å. For the simulations performed with the CHARMM force field, the Van der Waals interaction with the cations was modified to correct the excessive binding (NBFIX) [50,51]. All systems were first minimized by 2000 conjugate gradient steps, and then equilibrated by using a linear temperature gradient that heated the system from 0 to 300 K in 5 ns. All systems then were equilibrated for 10 ns at 300 K. Free molecular dynamics of all equilibrated systems were performed with a 2 fs integration time step using the RATTLE algorithm applied to all bonds. Constant temperature (300 K) was imposed by using Langevin dynamics [52], with a damping coefficient of 1.0 ps. A constant pressure of 1 atm was maintained with a Langevin piston dynamics [53], 200 fs decay period and 50 fs time constant. Each simulation was run up to 300 ns after the equilibration step (Table 1). Throughout the simulations, we monitored the order parameter, the area per lipid and the diffusion coefficient. We assumed that convergence was achieved, when the average and the standard deviation, computed on a 20 ns sliding window, reached a plateau (i.e. always within the first 100 ns of simulation). The analyses were performed on the last 150 ns of each simulation.

***A posteriori* validation of the cardiolipin parameterization**

NWChem was used to carry out a quantum mechanics/molecular mechanics (QM/MM) microiteration-based geometry optimization [54]. A set of 50 representative bilayer structures was extracted from CHARMM and AMBER all-atom MD simulations. Head-groups (the glycerol and both phosphate groups) were then randomly selected and optimized at the QM level. The defined QM and MM regions were treated with a DFT/B3LYP [55] functional using 6-31G* [40] basis set and Amber ff99SB force field [43], respectively. The DFT convergence was set at 10^{-7} Hartree. The MM charges were excluded for linked MM and QM atoms. A set of 300 cycles with 500 QM and 10'000 MM iterations was performed to reach a 10^{-6} convergence.

3.4.5 Supporting Information

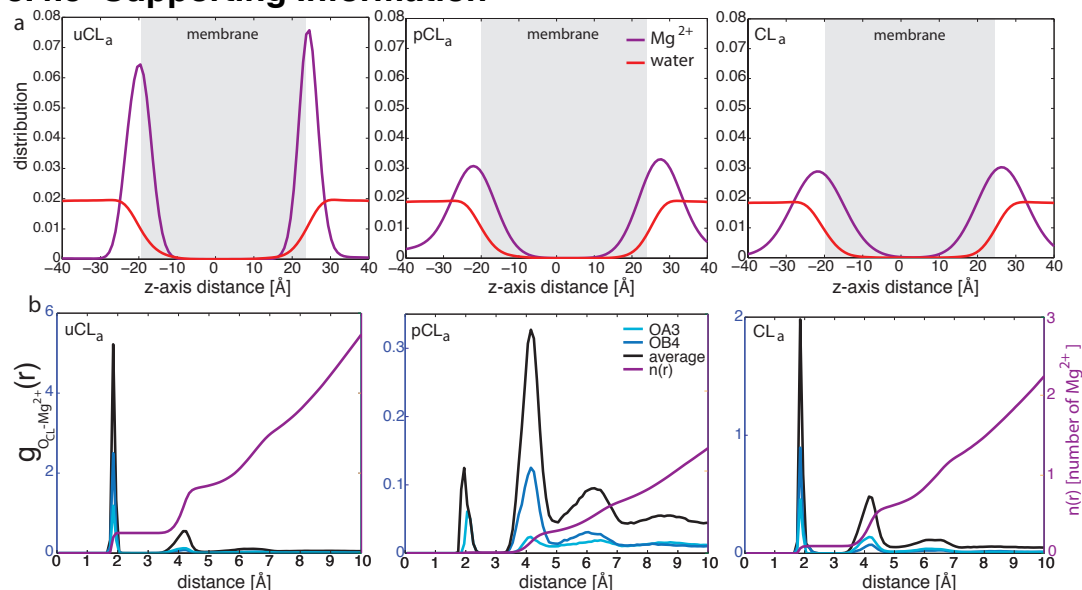


Figure S5: Cardioline interactions with divalent cations for CL models compatible with the AMBER force field. (a) Distribution of Mg^{2+} (purple) and water molecules (red) along the z-axis normal to the membrane surface. From left to right, data for the uCL, pCL and CL (mixed uCL/pCL) systems are reported when using the CHARMM force field. (b) Radial distribution function $g(r)$ and coordination number $n(r)$ (orange) of Mg^{2+} around phosphate oxygen atoms for simulations carried out with AMBER force field.

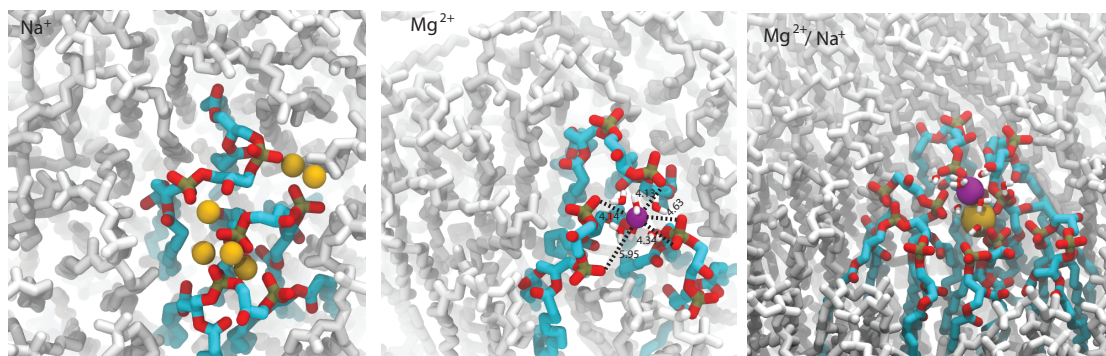


Figure S6: Coordination of cations around CL molecules. (Left) A single cardioline molecule interacts with up to three Na^+ cations (orange). The cation interacts directly with the phosphate group (see Figures 4 and 6 of the main text). (Middle) Four cardioline molecules (in cyan) trap a Mg^{2+} cation (purple) with its first solvation shell. The phosphate groups interact directly with the solvation shell are localized at ~ 4 Å from the cation, whereas the second phosphate group is localized at ~ 6 Å. (see Figures 5 and 6 of the main text). (Right) Na^+ cations bring the CL head group closer allowing it to trap a Mg^{2+} cation.

Table S1. PDB id used for data mining. PDB codes are reported below.

1E14, 1JGW, 1JGY, 1JGZ, 1K6L, 1K6N, 1KB9, 1KBY, 1KQF, 1KQG, 1M3X, 1NEK, 1NEN, 1OGV, 1OKC, 1P84, 1PP9, 1PPJ, 1QOV, 1RG5, 1RQK, 1RVJ, 1RY5, 1RZH, 1SQP, 1V54, 1V55, 1YF6, 2A06, 2ACZ, 2C3E, 2DYZ, 2DYS, 2EIJ, 2EIK, 2EIL, 2EIM, 2EIN, 2GNU, 2HG3, 2HG9, 2HH1, 2HHK, 2HIT, 2HJ6, 2J8C, 2J8D, 2JIY, 2JJ0, 2UWU, 2UWW, 2UX3, 2UX4, 2UX5, 2UXJ, 2UXK, 2WX5, 2YBB, 2YL4, 2ZXW, 3ABK, 3ABL, 3ABM, 3AG1, 3AG2, 3AG3, 3AG4, 3ASN, 3ASO, 3CWB, 3CX5, 3DSY, 3DTA, 3DTR, 3DTS, 3DU2, 3DU3, 3DUQ, 3H1H, 3H1I, 3H1J, 3H1K, 3H1L, 3I4D, 3L70, 3L71, 3L72, 3L73, 3L74, 3L75, 3MA7

3.4.6 References

1. Matsumoto K, Kusaka J, Nishibori A, Hara H (2006) Lipid domains in bacterial membranes. *Molecular Microbiology* 61: 1110-1117.
2. Lucero HA, Robbins PW (2004) Lipid rafts-protein association and the regulation of protein activity. *Archives of Biochemistry and Biophysics* 426: 208-224.
3. Wisniewska A, Draus J, Subczynski WK (2003) Is a fluid-mosaic model of biological membranes fully relevant? Studies on lipid organization in model and biological membranes. *Cellular and Molecular Biology Letters* 8: 147-160.
4. van Meer G, Voelker DR, Feigenson GW (2008) Membrane lipids: where they are and how they behave. *Nature Reviews Molecular Cell Biology* 9: 112-124.
5. Hoch FL (1992) Cardiolipins and Biomembrane Function. *Biochimica Et Biophysica Acta* 1113: 71-133.
6. Hoch FL (1998) Minireview: Cardiolipins and mitochondrial proton-selective leakage. *Journal of Bioenergetics and Biomembranes* 30: 511-532.
7. Kates M, Syz JY, Gosser D, Haines TH (1993) Ph-Dissociation Characteristics of Cardiolipin and Its 2'-Deoxy Analog. *Lipids* 28: 877-882.
8. Nichols-Smith S, Kuhl T (2005) Electrostatic interactions between model mitochondrial membranes. *Colloids Surf B Biointerfaces* 41: 121-127.
9. Pierucci O (1979) Phospholipid-Synthesis during the Cell-Division Cycle of *Escherichia-Coli*. *Journal of Bacteriology* 138: 453-460.
10. Houtkooper RH, Vaz FM (2008) Cardiolipin, the heart of mitochondrial metabolism. *Cellular and Molecular Life Sciences* 65: 2493-2506.
11. Egberts E, Marrink SJ, Berendsen HJC (1994) Molecular-Dynamics Simulation of a Phospholipid Membrane. *European Biophysics Journal with Biophysics Letters* 22: 423-436.
12. Heller H, Schaefer M, Schulten K (1993) Molecular-Dynamics Simulation of a Bilayer of 200 Lipids in the Gel and in the Liquid-Crystal Phases. *Journal of Physical Chemistry* 97: 8343-8360.
13. Bandyopadhyay S, Shelley JC, Klein ML (2001) Molecular dynamics study of the effect of surfactant on a biomembrane. *Journal of Physical Chemistry B* 105: 5979-5986.
14. Dahlberg M, Maliniak A (2008) Molecular dynamics simulations of cardiolipin bilayers. *Journal of Physical Chemistry B* 112: 11655-11663.
15. Dahlberg M (2007) Polymorphic phase behavior of cardiolipin derivatives studied by coarse-grained molecular dynamics. *Journal of Physical Chemistry B* 111: 7194-7200.
16. Aguayo D, Gonzalez-Nilo FD, Chipot CJ (2012) Insight into the properties of cardiolipin containing bilayers from molecular dynamics simulations, using a hybrid all-atom/united-atom force field. *Journal of Chemical Theory and Computation*.
17. Murzyn K, Rog T, Pasenkiewicz-Gierula M (2005) Phosphatidylethanolamine-phosphatidylglycerol bilayer as a model of the inner bacterial membrane. *Biophysical Journal* 88: 1091-1103.
18. Rog T, Martinez-Seara H, Munck N, Oresic M, Karttunen M, et al. (2009) Role of Cardiolipins in the Inner Mitochondrial Membrane: Insight Gained through Atom-Scale Simulations. *Journal of Physical Chemistry B* 113: 3413-3422.

19. Murzyn K, Pasenkiewicz-Gierula M (1999) Construction and optimisation of a computer model for a bacterial membrane. *Acta Biochimica Polonica* 46: 631-639.
20. Dahlberg M, Marini A, Mennucci B, Maliniak A (2010) Quantum chemical modeling of the cardiolipin headgroup. *J Phys Chem A* 114: 4375-4387.
21. Lewis RNAH, McElhaney RN (2009) The physicochemical properties of cardiolipin bilayers and cardiolipin-containing lipid membranes. *Biochimica Et Biophysica Acta-Biomembranes* 1788: 2069-2079.
22. Vermeer LS, de Groot BL, Reat V, Milon A, Czaplicki J (2007) Acyl chain order parameter profiles in phospholipid bilayers: computation from molecular dynamics simulations and comparison with ²H NMR experiments. *Eur Biophys J* 36: 919-931.
23. Lewis RNAH, Zweytick D, Pabst G, Lohner K, McElhaney RN (2007) Calorimetric, X-ray diffraction, and spectroscopic studies of the thermotropic phase behavior and organization of tetramyristoyl cardiolipin membranes. *Biophysical Journal* 92: 3166-3177.
24. Shinoda W, Okazaki S (1998) A Voronoi analysis of lipid area fluctuation in a bilayer. *Journal of Chemical Physics* 109: 1517-1521.
25. Bockmann RA, Hac A, Heimburg T, Grubmuller H (2003) Effect of sodium chloride on a lipid bilayer. *Biophys J* 85: 1647-1655.
26. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceeding of the National Academy of Sciences of the USA* 98: 10037-10041.
27. Dickson CJ, Rosso L, Betz RM, Walker RC, Gould IR (2012) GAFFlipid: a General Amber Force Field for the accurate molecular dynamics simulation of phospholipid. *Soft Matter* 8: 9617-9627.
28. Khalifat N, Fournier JB, Angelova MI, Puff N (2011) Lipid packing variations induced by pH in cardiolipin-containing bilayers: the driving force for the cristae-like shape instability. *Biochim Biophys Acta* 1808: 2724-2733.
29. Haiens TH, Dencher NA (2002) Cardiolipin: a proton trap for oxidative phosphorylation. *Febs Letters* 528: 35-39.
30. Lange C, Nett JH, Trumpower BL, Hunte C (2001) Specific roles of protein-phospholipid interactions in the yeast cytochrome bc₁ complex structure. *EMBO J* 20: 6591-6600.
31. Tatulian SA (1987) Binding of Alkaline-Earth Metal-Cations and Some Anions to Phosphatidylcholine Liposomes. *European Journal of Biochemistry* 170: 413-420.
32. Huster D, Arnold K, Gawrisch K (2000) Strength of Ca²⁺ binding to retinal lipid membranes: Consequences for lipid organization. *Biophysical Journal* 78: 3011-3018.
33. Garidel P, Blume A, Hubner W (2000) A Fourier transform infrared spectroscopic study of the interaction of alkaline earth cations with the negatively charged phospholipid 1,2-dimyristoyl-sn-glycero-3-phosphoglycerol. *Biochimica Et Biophysica Acta-Biomembranes* 1466: 245-259.
34. Macdonald PM, Seelig J (1987) Calcium-Binding to Mixed Cardiolipin Phosphatidylcholine Bilayers as Studied by Deuterium Nuclear-Magnetic-Resonance. *Biochemistry* 26: 6292-6298.

35. Koppelman CM, Den Blaauwen T, Duursma MC, Heeren RMA, Nanninga N (2001) *Escherichia coli* minicell membranes are enriched in cardiolipin. *Journal of Bacteriology* 183: 6144-6147.
36. Som A, Yang L, Wong GC, Tew GN (2009) Divalent metal ion triggered activity of a synthetic antimicrobial in cardiolipin membranes. *Journal of the American Chemical Society* 131: 15102-15103.
37. Schroedinger L (2010) Maestro, version 8.5. Schroedinger, LLC, New York.
38. Jorgensen WL, Maxwell DS, TiradoRives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* 118: 11225-11236.
39. Frisch M, Trucks G, Schlegel H, Scuseria G, Robb M, et al. (2009) Gaussian-09, Rev. A. 1; Gaussian. Inc: Wallingford, CT.
40. Raghavachari K (2000) Perspective on "Density functional thermochemistry. III. The role of exact exchange" - Becke AD (1993) *J Chem Phys* 98:5648-52. *Theoretical Chemistry Accounts* 103: 361-363.
41. Mayne CG, Tajkhorshid E, Gumbart J (2012) Force Field Toolkit Plugin. www.kisu.edu/Research/vmd/plugins/fftk.
42. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, et al. (2010) Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *Journal of Physical Chemistry B* 114: 7830-7843.
43. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, et al. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins-Structure Function and Bioinformatics* 65: 712-725.
44. Tessier MB, DeMarco ML, Yongye AB, Woods RJ (2008) Extension of the GLYCAM06 biomolecular force field to lipids, lipid bilayers and glycolipids. *Molecular Simulation* 34: 349-363.
45. Tunaitis E, Cronan JE, Jr. (1973) Characterization of the cardiolipin synthetase activity of *Escherichia coli* envelopes. *Arch Biochem Biophys* 155: 420-427.
46. Gold VA, Robson A, Clarke AR, Collinson I (2007) Allosteric regulation of SecA: magnesium-mediated control of conformation and activity. *J Biol Chem* 282: 17424-17432.
47. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 26: 1781-1802.
48. Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr., Nilsson L, Petrella RJ, et al. (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30: 1545-1614.
49. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 79: 926.
50. Mayaan E, Moser A, MacKerell AD, Jr., York DM (2007) CHARMM force field parameters for simulation of reactive intermediates in native and thio-substituted ribozymes. *J Comput Chem* 28: 495-507.
51. Noskov SY, Berneche S, Roux B (2004) Control of ion selectivity in potassium channels by electrostatic and dynamic properties of carbonyl ligands. *Nature* 431: 830-834.

52. Brunger A, Brooks CL (1984) Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chemical Physics Letters* 105: 495-500.
53. Feller SE, Zhang Y, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation: the Langevin piston method. *The Journal of Chemical Physics* 103: 4613.
54. Valiev M, Bylaska EJ, Govind N, Kowalski K, Straatsma TP, et al. (2010) NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Computer Physics Communications* 181: 1477-1489.
55. Wadt WR, Hay PJ (1985) Ab initio effective core potentials for molecular calculations. Potentials for main group elements Na to Bi. *The Journal of Chemical Physics* 82: 284.

3.5 Lipid Builder for membrane models

The physicochemical properties of the different lipid species are essential for the structural and functional organization of the biological membrane. They play a key role in the segregation of membrane domains, the formation of non-bilayer structures, and the interactions with proteins. The lipid head groups are not the only important determinants of a membrane. The nature of the lipid acyl chains strongly influences membrane properties. They define the membrane thickness and fluidity, and consequently the orientation of the membrane proteins. Unsaturation changes the packing of the lipids and creates defects in the bilayer, which affect the insertion of molecules and peptides. Due to the increase in computational power, it is now possible to simulate large membrane systems composed of a complex mixture of lipids.

To facilitate the creation of templates and topology files, a VMD plug-in based on the CHARMM force field was implemented. Based on the SMILE structure of the acyl chains, the creation of the corresponding phospholipid is fully automatized. First, the lipid's topology is created by combining the selected phospholipid head groups, extracted from a built-in library of structures and the provided acyl chains. Since the acyl chains have been parameterized based on a "plug and play" philosophy in the CHARMM force field, the acyl chains are generated by linking a series of alkanes. Three different classes of hydrocarbons have been defined: saturated, unsaturated and cyclopropane. We used the refined QM-based parameters to better describe the cyclopropane ring [1]. This cyclic moiety has frequently been neglected in the MD simulations of bacterial membranes. They are however present in many species of bacteria and increase the lipid surface area. Finally, the psfgen algorithm and the generated topology file were combined to build the template PDB and PSF files of the phospholipids [2].

Using this plug-in, a realistic bacterial membrane composed of the 20 main species of *E. coli* phospholipids was assembled. The bilayer is a model of the pole region characterized by a cardiolipin rich domain. It is composed of 50% cardiolipins, 35% PE and 15% PG (Figure 18). We combined the results of two mass spectrometry experiments to define the acyl chain composition of the lipids [3,4]. The analysis of this membrane model using MD will be important for better understanding the biophysical properties of the bacterial membrane and for guiding the rational design of novel antibiotics mimicking the AMP mechanism of action.

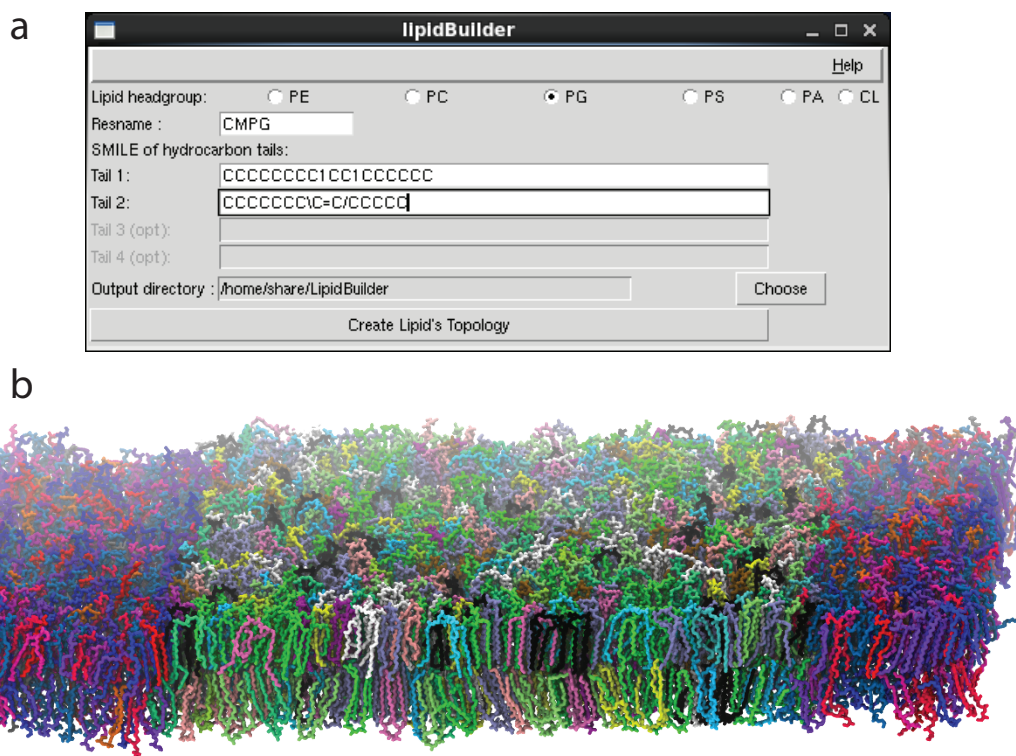


Figure 18: LipidBuilder. (a) The graphical user interface of LipidBuilder, and (b) the membrane bilayer of a cardiolipin-rich domain. Each lipid type is displayed in a different color.

3.5.1 References

1. Pandit KR, Klauda JB (2012) Membrane models of E-coli containing cyclic moieties in the aliphatic lipid chain. *Biochimica et Biophysica Acta-Biomembranes* 1818: 1205-1210.
2. Gullingsrud J, Saam J, Phillips J (2006) psfgen User's Guide. Urbana 51: 61801.
3. Ishida Y, Kitagawa K, Nakayama A, Ohtani H (2006) Complementary analysis of lipids in whole bacteria cells by thermally assisted hydrolysis and methylation-GC and MALDI-MS combined with on-probe sample pretreatment. *Journal of Analytical and Applied Pyrolysis* 77: 116-120.
4. Hsu FF, Turk J (2006) Characterization of cardiolipin from *Escherichia coli* by electrospray ionization with multiple stage quadrupole ion-trap mass spectrometric analysis of $[M - 2H + Na]^-$ ions. *J Am Soc Mass Spectrom* 17: 420-429.

3.6 Conclusions and Perspectives

The pharmacological use of antimicrobial peptides represents a promising alternative to conventional drugs. Their use and future development however require a better knowledge of bacterial sensory systems and membrane response. Therefore, this thesis work focused on producing accurate models of the bacterial chemosensor PhoQ and membrane constituents.

PhoPQ TCS senses the host environment via antimicrobial peptides and acidic pH. Since two-component systems appear to be absent in the animal cell, they also represent a promising target for the development of synthetic antimicrobial drugs. A significant obstacle in the understanding of signal transduction is the lack of a full atomistic model for the signaling TCS machine. Therefore, the assembly of an atomistic model of the PhoQ TCS was carried out in this thesis project. The structures of certain domains of the PhoQ have already been solved, and others were produced by comparative modeling here. A structural model of the transmembrane region is crucial for understanding signal transduction across the membrane. However, no structure or homology-model is available and its modeling represents a major challenge. The final PhoQ model produced by linking these different sub-domains (Figure 10) supports a transduction mechanism consistent with available mutagenesis studies and could help to rationalize the sensing of antimicrobial peptides and the design of antimicrobial drugs. PhoQ plays a key role in the sensing the host environment and the regulation of the transcription of genes promoting bacterial resistance to antimicrobial peptides. Therefore the inhibition of PhoQ could impair the virulence of some species of bacteria (*e.g. Salmonella typhimurium*) and improve the efficiency of current antimicrobial peptides.

Antibiotic selectivity is based on fundamental biochemical differences between microbial and mammalian cells. Membrane composition appears to highly affect the selectivity of AMPs. Models for a realistic bacterial membrane whose composition is as close as possible to cellular conditions are therefore desirable. In this context, cardiolipin species, which are key components of the bacterial membrane, were successfully parameterized and characterized. Using these models, the investigation of a large system representative of the bacterial membrane at the pole region is currently being carried out (Figure 18), with the hope of better analyzing and understanding the distinctive anionic features relevant for antibiotic resistance.

Amyloid Precursor Protein

4.1 The flexibility of the transmembrane domain of the amyloid precursor protein

Submitted as the following paper:

“The amyloid precursor protein APP-TM is highly flexible and adapts to the lipid environment”,

Lemmin T., Dimitrov M., Fraering P. C., and Dal Peraro M.

4.1.1 Introduction

The aggregation of amyloid- β ($A\beta$) peptides and the formation of extracellular $A\beta$ plaques in brain tissue have been associated to the neuropathological process of Alzheimer’s disease (AD). $A\beta$ is a 38 – 42 amino acid long peptide ($A\beta_{38-42}$) contained within the C-terminal fragment of the Amyloid Precursor Protein (APP), a type-I transmembrane glycoprotein. $A\beta$ is produced by the sequential cleavage of APP by β - and γ -secretases. The β -secretase first cleaves the extracellular domain of APP, producing a 99 amino acid long transmembrane protein, referred to as APP-C99. Then, the proteolytic cleavage of APP-C99 by γ -secretase occurs within the transmembrane domain, releasing $A\beta$ peptides.

The neurotoxicity of $A\beta$ depends on the peptide length and therefore on the position within APP-C99 of the γ -secretase cleavage site. In Alzheimer’s patients, the ratio of $A\beta_{42}$ versus $A\beta_{40}$ is significantly increased. Effort has therefore been invested into characterizing the structure and dynamics of APP-C99 and how this substrate would interact with γ -secretase. Mutations in or adjacent to the transmembrane region have a drastic effect on cleavage at the γ -site. More than half of the mutations in APP causing early-onset familial AD are found in its transmembrane domain (TM) (3). Moreover, the interaction of the APP-TM with the membrane environment, in particular with cholesterol, Cu^{2+} and Zn^{2+} metal ions, alter the structural and dynamical properties of the protein (4, 5).

Until recently, only a theoretical structure existed for the APP transmembrane domain. Miyashita *et al.* (6) used molecular dynamics simulation (MD) in implicit solvent to build and study a first atomistic model of the APP transmembrane domain. They observed that G708 and G709 play an important role in the flexibility of the helix. This mechanism can stabilize the TM domain in the membrane, allowing it to

adapt to the fluctuations of the membrane thickness. These residues are close to the γ -site and may directly influence cleavage. The APP-TM contains two GXXXG motifs (Figure S7), which are believed to mediate dimerization in transmembrane proteins (7). Mutations of the GXXXG motif also lead to a drastic decrease in the secretion of A β_{42} (8-10). Molecular dynamics simulations in implicit solvent showed that these mutations shifted the position of the γ -site in the membrane (11). Recently, two NMR structural ensembles have been solved. Nadezhdin *et al.* (2) determined the structure of APP-TM (Q686 – K726) in dodecylphosphocholine (DPC) detergent micelles (PDB code: 2LLM, Figure 1a). The structure features a short N-terminus juxta-membrane helix (K687 – D694) and a 24 amino acid transmembrane helix (G700 – L723). A minor bend (~ 16 degrees) is observed at residues G708 – G709. These residues, located in the middle of the TM, can act as a hinge to compensate for the fluctuation of the membrane thickness.

Barrett *et al.* (1) solved the NMR structure for the helical domain of APP-C99 (V683 – Y728) in lysomyristoylphosphatidylglycerol (LMPG) detergent micelles (PDB code: 2LP1, Figure 19a). The N- and C- termini are mostly unstructured, except for a short C-terminus helix, whereas the transmembrane domain is characterized by a highly curved α -helix (~ 115 degrees at G708 – G709). To validate the biological relevance of this kink, double electron-electron resonance (DEER) spectroscopy was used to estimate the distance between the ends of the TM domain in lipid bilayers. Methanethiosulfonate (MTSL) spin labels were attached to residues G700 and L723, mutated beforehand to cysteines. The distance measured in the micelles and lipid vesicles were nearly identical, thus providing indirect evidence that APP would also be curved in lipid bilayers.

In order to explain the existence of these divergent structures (Figure 19a) and to obtain detailed insights into the dynamic properties of APPs, we used molecular dynamics simulations to understand the effect of the different experimental conditions on the APP transmembrane conformations. We show that the kink observed in the APP NMR structures is largely reduced when inserted into a more physiological lipid bilayer environment. The relative conformational space of APP is strongly influenced by the size of the micelles and by perturbations introduced by specific mutations and spin labels. Taken together our results support a conformation for the APP transmembrane domain that is mainly populated by straight α -helical structures when embedded in a membrane bilayer. This conformation however shows a significant flexibility at the G708 – G709 region, which explains the multiple conformations that have been observed. This flexibility may have functional implications for γ -secretase presentation during proteolytic processing.

4.1.2 Results

We carried out an extensive set of MD simulations of APP-TM structures considering their embedment into a palmitoyloleoylphosphatidylcholine (POPC) lipid bilayer and LMPG micelles. We compared and validated our results with available experimental data (namely, NMR and EPR (1)). We also tested the effect of the experimental protocols; such as those produced by spin labeling of APPs. MD simulations of proteins in the membrane might require a long simulation time to reach equilibrium. We therefore used enhanced sampling techniques such as metadynamics to explore

the dynamic determinants of the APP-TM domain and to reconstruct the corresponding free energy landscape under relevant external conditions (12).

The APP-TM domain has a predominantly straight helical conformation in a membrane bilayer.

We used molecular dynamics (MD) to study the structural and dynamic properties of the APP transmembrane domain. We considered the two available NMR structural ensembles (PDB codes 2LLM and 2LP1), referred to hereinafter as APP_{2LP1} and APP_{2LLM}, respectively. APP_{2LLM} is solved as a mostly straight α -helix with a small concavity near G708 – G709, whereas the APP_{2LP1} structure is characterized by an important kink at this region. To characterize these structures in a physiological-like environment, we inserted them into a 60 x 60 Å² membrane bilayer formed by phosphatidylcholine (POPC), which is a major lipid constituent of the synaptic plasma membrane (13). The MD simulations were carried out under physiological-like conditions (e.g., pH ~7, 1 atm and 300 K, Figure 19b).

When inserted into the POPC bilayer, the secondary structure of both APP NMR structures fluctuates only marginally and preserves an α -helical conformation. The transmembrane domain of APP_{2LLM} remains very close to the initial NMR structure (RMSD 1.86 \pm 0.19 Å). The important bend observed in the APP_{2LP1} initial structure straightens within the first 20 ns of simulation and remains straight throughout the rest of the simulation (>300 ns, Table S1). The final structures for the transmembrane domain, starting from both initial conformations, are nearly identical (RMSD 1.26 \pm 0.14 Å) and are characterized by a straight TM α -helix with a small kink at the paired residues G708 – G709 (Figure 19b). The APP_{2LLM} TM helix tilts slightly less with respect to the membrane surface compared to APP_{2LP1} ($\sim 20 \pm 7$ and $\sim 36 \pm 5$ degrees, respectively). We also observe a different behavior of the N-terminus domain. Consistent with the NMR observations and theoretical models, the APP_{2LP1} N-terminus amphiphilic helix (Q686 – N698) lies on the membrane surface, and a lysine belt (K687, K699, K724 – 726) anchors the TM and N-terminus helices, respectively, in and onto the membrane. In the APP_{2LLM}, an extra turn forms at the N-terminus of the TM domain (S697 – K699), hindering the interaction of the N-terminus helix with the membrane surface. This is however consistent with the NMR ensembles, where this terminal helix is flexible and explored a broad spectrum of conformation (angles ranging from 75° – 180° with respect to the TM domain).

The water molecules interact mostly with residues close to the membrane interface (G700, A701, L720 and L721), and a small solvation is observed in the core of the TM domain, where water molecules interact with G708 or G709 (Figure 19c). The water molecules access the TM core from the N-terminus by interacting with the G₂₉XXXGXXXG₃₇, or from the C-terminus by interacting with the residue T719 side chain (Figure S1).

In order to characterize the free energy landscape associated with the helix bending in a membrane bilayer, we defined the two dihedral angles ϕ and ψ , respectively connecting the backbone nitrogen atoms of G708 to G709 and G709 to V710. Consistent with the unbiased MD simulations, the metadynamics trajectories showed that the POPC bilayer presence strongly confines the APP-TM structure to a well-defined straight helical conformation (minimum at [-35°; -43.5°], Figure 19d). The free

energy penalty needed to explore bent conformations similar to the APP_{2LP1} NMR structures is as high as 6-7 kcal/mol, suggesting that in the membrane environment, a straight conformation is more populated than the curved one.

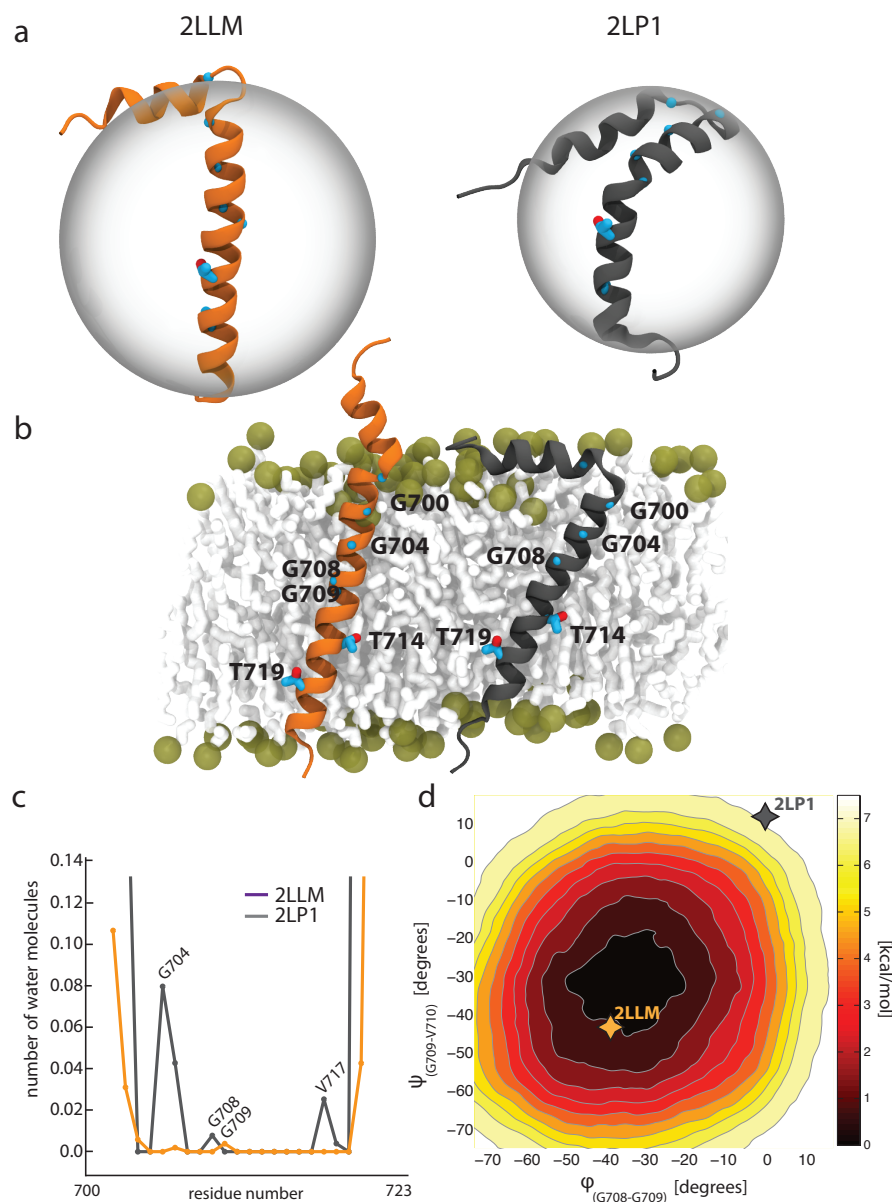


Figure 19: APP-TM under physiological conditions. (a) The NMR structures of the APP-TM domain (PDB code: 2LLM (2) and 2LP1 (1)). The hypothetical micelle sizes inscribing the protein are displayed with transparent spheres and are based on the experimental setup reported in (1, 2), (b) Representative MD snapshots of the NMR 2LLM (left) and 2LP1 (right) structures after 200 ns in a POPC bilayer, (c) The relative level of solvation of each APP-TM residue for the two systems in the bilayer and (d) Free energy landscape associated to ϕ and ψ dihedral angles connecting the nitrogen atoms of G708 to G709 and G709 to V710, respectively, as calculated by metadynamics MD. The average NMR conformations relative to

the straight 2LLM [-41°; -46°] and bent 2LP1 [-1°; 13°] ensembles are reported by green asterisks.

The APP-TM domain explores α -helical kinked conformations in detergent micelles.

Since both NMR structural ensembles were solved in detergent micelles and not in a lipid bilayer, we used MD simulations to study the behavior of APP when embedded in micelles having properties mimicking the experimental setup. The APP_{2LP1} NMR ensemble was used as starting structures in order to investigate if its tertiary structure could be stabilized by the micelle. The micelles inscribing the APP_{2LP1} and APP_{2LLM} NMR structures are reported to have different sizes (Figure 19a) due to the different lipid detergents used in the experiments. Two systems containing the juxta-membrane and TM helices (residues V683–Y728) embedded into different micelles were therefore assembled. As in Barrett *et al.* (1), a 35-Å diameter micelle was considered for the first system (M35), whereas a larger micelle (i.e. 45 Å diameter) was used in the second system (M45) to study the dependence of the APP conformation on micelle size. Each micelle was first equilibrated separately in a water box. The micelle assembly converged and remained stable after 50 ns. The protein was then inserted into the micelle and the simulations were extended for at least an extra 150 ns. Since the NMR spectra were measured at 45 °C, we performed simulations under physiological conditions (300 K) and at a higher temperature (318 K) consistent with the experimental setup (Table S1).

The M35 micelle severely affects the structure of APP. The size of this micelle is too small to fully accommodate the TM helix, thus exposing the helical interface that encloses the G₂₉XXXGXXXG₃₇ motif to the solvent (Figure 20a). The T714 and T719 hydroxyl groups and the G700, G704, G708 and V711 backbone atoms participate the most in the interaction with water molecules (Figure 20b). The TM helix is highly curved, kinking either between residues G708 and G709, which is consistent with the NMR, or more rarely between residues G709 and V710. The larger M45 micelle can better accommodate the transmembrane domain (Figure 20a). The structure alternates between a curved and straight structure, with a preference for the latter one. Similar to M35, but to a lesser extent, the threonine side chains and glycine backbone interact with water molecules (Figure 20b). The solvation of the TM domain therefore seems to be the major driving force producing a kink in the APP-TM domain. Since only one interface is exposed to the solvent, solvation is observed with the same periodicity as the helical motif (Figure 20b). Consistent with the NMR structure, the amphiphilic N-terminus helix lies on the surface of the micelles. Heating of the system to 318 K, as in the experimental setup, only marginally affected the APP structure. The ensemble of conformers remained comparable to the one at 300 K (population RMSD 2.86 ± 0.69 Å and 3.01 ± 0.39 Å, respectively).

The NMR APP_{2LP1} structural ensemble however is characterized by a single conformation of the TM backbone. In order to explain the discrepancy between our observations and the NMR results, we decided to back-compute the residual dipolar couplings (RDCs). The RDCs provide long-range structural information and can be used to estimate the fold of proteins. They have been shown to be in good agreement with the results of MD simulations (14). Since the exact characteristics

and composition of the NMR sample cannot be defined, we used a genetic algorithm (GA) to select an ensemble of conformers that would best fit the experimental RDCs. A set of 3000 structures was first extracted from the MD simulations performed in both M35 and M45 (similar results were obtained by a separate analysis of the M35 and M45 systems). For each structure, the backbone N–H vector was defined and the alignment tensor was computed using singular value decomposition (SVD). Based on several runs of the GA, on the average, a set of 500 structures was able to best fit the experimental RDCs. The correlation of these back-calculated RDCs extracted from the MD trajectories was comparable to the one calculated for the NMR ensemble (Q-factor 0.23 and 0.31, respectively, Figure S8a). The selected structures were finally grouped into 33 structures composed of straight and bent helices obtained from the M35 and M45 simulations (Figure S8b). This provided a solid indication that APP is able to explore a larger and more heterogeneous set of conformations in the micelle environment, as documented by MD simulations.

Finally, metadynamics simulations in M35 and M45 micelles were used to more exhaustively explore the conformational space of APP (Figure 20c). The free energy landscape for both micelles validated the unbiased simulations where the structures inserted into a micelle could freely explore the bent and straight conformations on a multi-nanosecond timescale. The barrier to pass from the straight helix minimum (slightly more populated) to a bent structure is in fact in the order of ~ 3 kcal/mol, and it is thus accessible via thermal fluctuations (Figure 20c). Consistent with our simulations, the conformational space associated with this transition is expected to be sampled on a $\sim 10^2$ ns time scale.

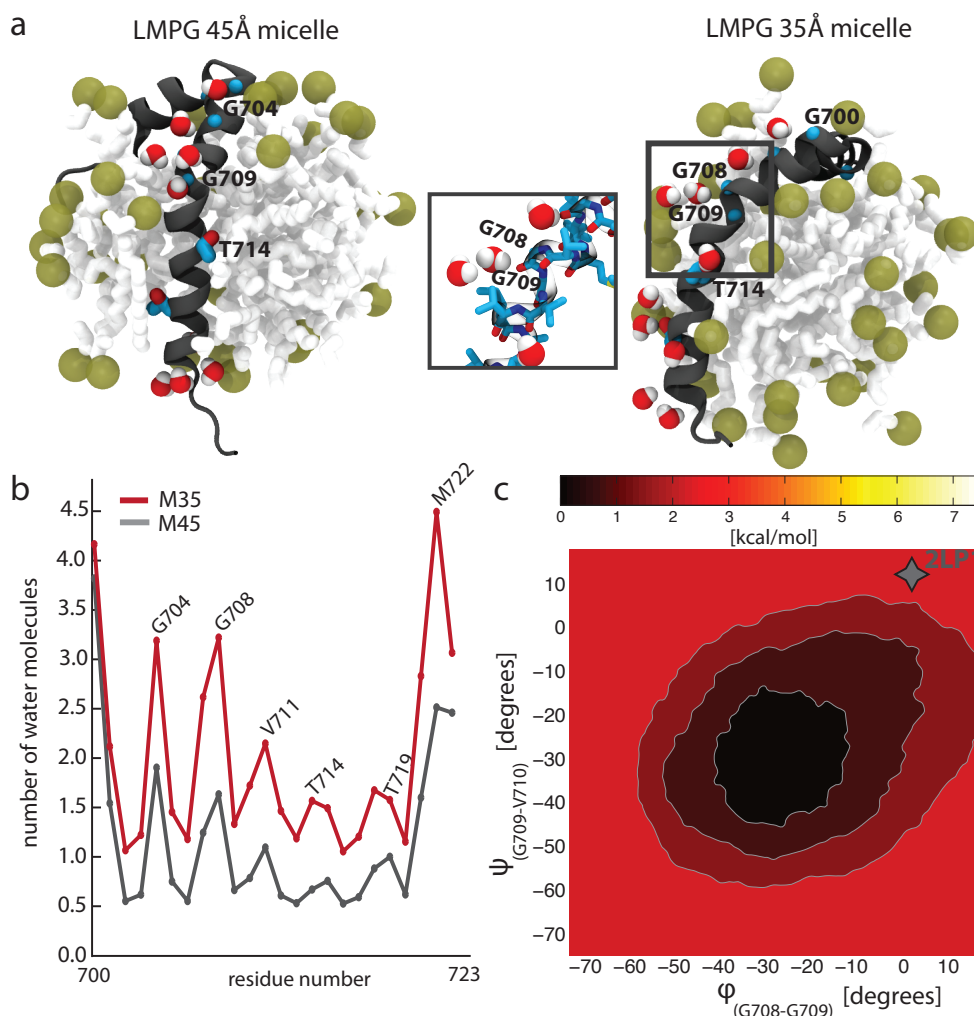


Figure 20: APP-TM in a water-micelle environment. (a) Representative MD snapshots of the APP2LP1 TM domain embedded into LMPG micelles of different sizes (35 and 45 Å, called respectively, M35 and M45) are shown. Glycine and threonine residues shown in blue are mainly exposed to the solvent. A focus of the G708-G709 region exposed to water molecules is reported in the inset, (b) The distribution of the water molecules solvating the TM-APP tract is reported for the M45 and M35 systems and (c) Free energy landscape associated to ϕ and ψ dihedral angles as calculated by metadynamics MD on the M45 system (similar results obtained for M35 are shown in **Figure S9**). The green asterisk indicates the average NMR conformation relative to the 2LP1 ensembles.

The APP-TM segment responds to structural perturbations by exploiting its flexibility

Electron paramagnetic resonance (EPR) power saturation methods can be used to measure the membrane depth of a protein. By monitoring the collision rate of a spin label with a paramagnetic relaxation agent, the solvent accessibility of the label can be estimated. Methanethiosulfonate (MTSL) is a commonly used nitroxide spin label, covalently bonded to genetically engineered cysteines in the protein (Figure 21a). The depth parameter is determined by the ratio of the collision rate of two different

probes (e.g. oxygen and $\text{Cr}(\text{C}_2\text{O}_4)^{3-}$) with the label and can be defined as a function of the distance of the label from the phosphate group (15, 16). EPR double electron-electron resonance (DEER) has become a popular method for measuring the distance distributions between paramagnetic spin labels in proteins (17) when two MTSL labels are bonded to cysteines. The DEER technique has the advantage that it neither requires the crystallization of the protein complex nor is it limited by the complex size. Distances in the range of 18 Å to 60 Å can be measured in biomolecules (18-20).

We used the EPR data reported by Barrett *et al.* (1) to clarify and rationalize our MD observations for the APP structure in the membrane bilayer. We also performed several point mutations in the APP-TM in order to characterize their impact on the tertiary structure. Two mutation sites (G700 and L723) used for the insertion of the MTSL spin labels in the DEER EPR were considered. First the single and the double mutants (G700C, L723C and G700C/L723C) were inserted into a POPC membrane and simulated under physiological conditions. The mutation on the G700 site significantly affected the structure and the dynamics of the APP. The position of the N-terminus helix on the membrane significantly differed from the wild-type (Figure S10). This mutation changes the flexibility of the loop connecting the juxta-membrane helix to the TM domain. The solvation of the APP-TM core is therefore also affected (Figure S11). In contrast, the L723C mutation did not significantly alter the structure of the APP. The tilting and position of the juxta-membrane helix remained similar to the WT.

For all three mutants, we attached a MTSL spin label to each cysteine (G700CM, L723CM and G700CM/L723CM). Again, the G700 mutations perturb the APP structure the most, with the TM helix remaining strongly curved and the N-terminus helix rearranging itself on the top of the membrane. The helix started to slowly straighten after 130 ns of simulation (Figure S10). The G700CM/L723CM system was therefore extended to 300 ns during which time the helix remained fully straight. As observed previously, the L723CM single mutant only marginally affected the APP structure, with the transmembrane domain remaining straight, as in the wild-type conformation.

The depth parameter of APP inserted in the lipid bilayer was computed for the straight and curved structures of the G700CM/L723CM mutant, respectively (Figure 21b). A representative population composed of 500 structures (over 50 ns) was extracted for each conformation. The power saturation profile of the straight structure better matches the experimental values than the curved one (correlation coefficient 0.96 and 0.88, respectively). Consistent with the experimental data, the most buried residues are located at around V711 for the straight structure. This maximum shifts from V711 to I716 for the bent structure. The profile of the bent helix for the flexible domain (G700 – G709) better matches the experimental values. As observed during simulations, this region is easily perturbed by the experimental protocol and bent when spin labels were attached.

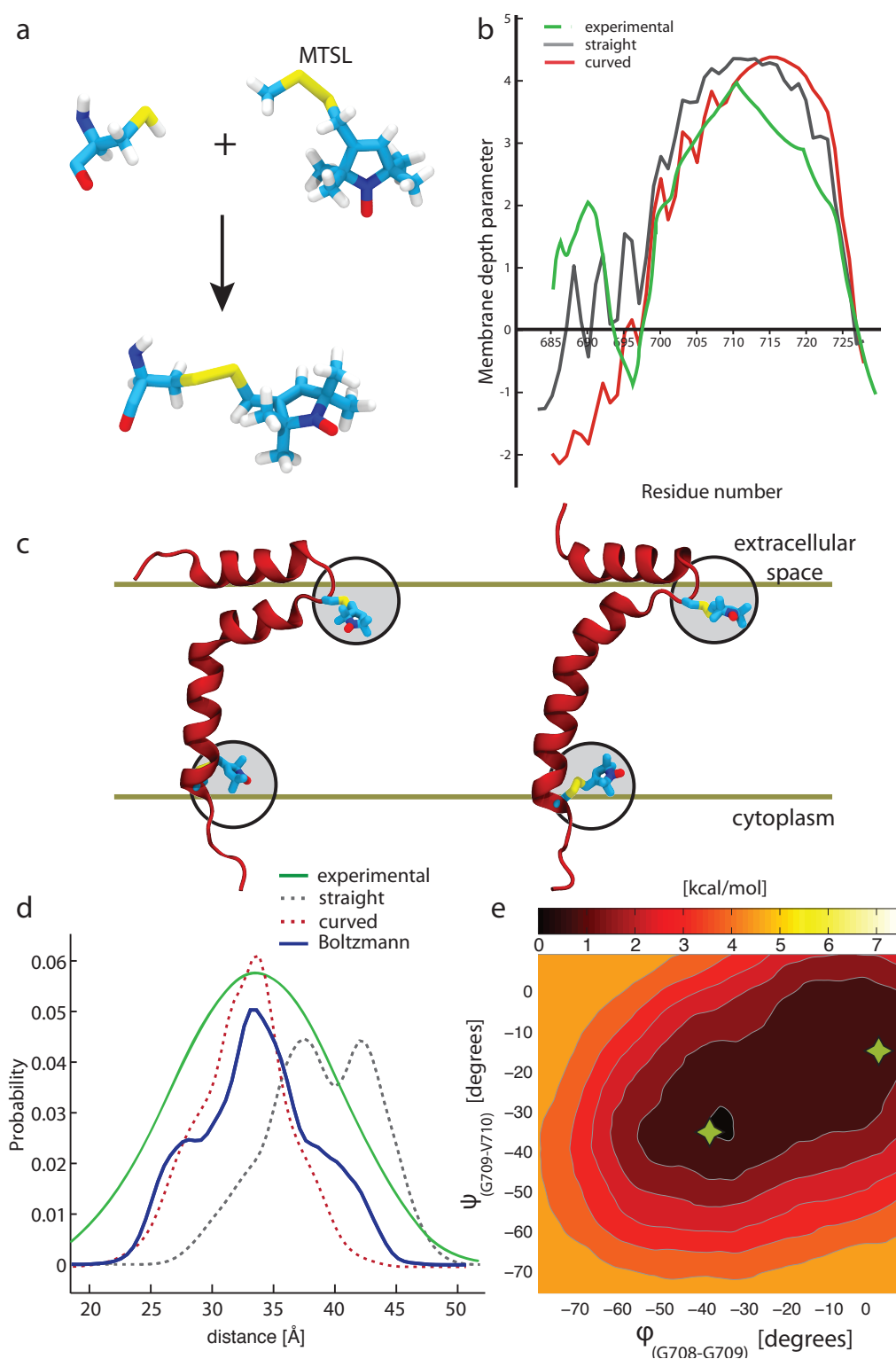


Figure 21: Effects of spin-labeled probes in EPR experiments. (a) Cysteine thiol side chain reacts with the MTSL spin label and produces a spin-labeled side chain, (b) The membrane depth parameter measured experimentally (1) (shown in green) is compared to the corresponding value computed from MD simulations for the straight and curved spin-labeled

TM conformation in black and red, respectively. Positive values indicate burial into the membrane, (c) Representative MD snapshots of the MTSL spin labels attached to the curved (left) and straight (right) APP (a solid line indicates the average position of the membrane bilayer). The black circles outline the possible conformational space accessible to the MTSL in solution; due to the hydrophobic nature of the MTSL tag, the membrane restricts this conformational space (highlighted in grey), (d) The MTSL distance distributions are compared as reported from DEER experiments (1) (in green), as predicted by using MMM for the MD ensembles of curved and straight conformations (in red and black dashed lines, other predictions are reported in Table S2), and as measured from unbiased and metadynamics MD simulations of the G700CM/L723CM system (in blue) and (e) Free energy landscape associated to ϕ and ψ dihedral angles as calculated by metadynamics MD for the G700CM/L723CM system. The average curved $[0^\circ; -20^\circ]$ and straight $[-40^\circ; -37^\circ]$ conformations shown in (c) are indicated by green asterisks.

The predicted distance distribution calculated using three different DEER prediction algorithms (MMM (21), mtsslWizard (22), and Pronox (23)) were compared for the straight and curved helical conformations extracted from MD simulations. All three algorithms predicted a similar distance and on the average, the distance for a bent helix was 31.3 ± 3.7 Å, and 38.9 ± 4.4 Å for the straight helix (Figure 21d for MMM prediction, and Table S2). When compared with the experimental distribution, these predictions support a more curved helical conformation. However, the straight conformation cannot be excluded, since these algorithms are parameterized to predict the distance of proteins in solution (17). The presence of the membrane would limit the flexibility of the probes and, most importantly, can strongly trap their hydrophobic regions. We therefore decided to extract the position of the N–O bond from the MTSL tags during MD simulations and compute the corresponding distance distribution (Figure S12). As expected, the distributions systematically shift towards shorter distances in comparison to DEER prediction algorithms. This difference is well pronounced for the straight helix distribution, and is comparable to the bent helix distribution. Due to the hydrophobic nature of the spin labels, they are buried deep into the membrane, thus limiting the conformational space of their rotational degrees of freedom (Figure S13). Furthermore, the presence of this anchor allows the TM helix to explore a bent conformation in the membrane bilayer.

Metadynamics simulations confirmed that by adding MTSL labels to APP at positions 700 and 723, the structure acquired an increased flexibility compared to the wild-type, and was able to explore more curved structures, as was observed in the micelle environment. A straight conformer is however slightly more preferred by ~ 1 kcal/mol over a bent conformation (Figure 21e). This is consistent with the observation in free dynamics, where the bent helix straightens. Since the ϕ dihedral explores more freely its conformational space, G708 may confer more flexibility to the APP-TM domain. Based on the free energy landscape, we built an ensemble of conformers that follow a Boltzmann distribution. The distance distribution computed for this ensemble allows to better account for the broad distribution observed experimentally (Figure 21d), and strongly hints to a heterogeneous population alternating between straight and bent APP conformations.

4.1.3 Discussion

The alteration in the proteolytic cleavage of the APP transmembrane domain has been linked to the pathogenesis of Alzheimer's disease. The absence of direct structural information on the APP/ γ -secretase complex currently hinders a complete understanding of the enzymatic mechanism and its pathogenic implications. Recently, two structures of the APP transmembrane domain were solved using NMR (PDB code: 2LLM and 2LP1), thus providing new insights into the intrinsic properties of APP. In the first structural ensemble, the APP-TM is mainly defined by a straight helical conformation, whereas an important bend is observed at residues G708 – G709 for the second reported NMR ensemble. Since these structural ensembles have been solved in detergent micelles, they are not fully representative of the APP embedded into its natural environment. In this study, we used molecular dynamics simulations coupled to NMR and EPR data to dissect the structural and dynamical properties of the transmembrane domain of APP when inserted into micelles and into a membrane bilayer that is more representative of the neuronal synaptic envelop.

We found that once inserted into a lipid bilayer, the transmembrane domain of the APP mainly adopts a straight helix conformation that is consistent with the 2LLM structural ensemble. The TM domain is characterized by a dynamic structure, in which the paired residues G708–G709 confer flexibility to the transmembrane domain and allow it to adapt to the fluctuations of the membrane thickness. To rationalize the origin of the discrepancy between the structure of APP in a bilayer and in micelles, we modeled as closely as possible the experimental conditions. We first inserted APP into micelles of different sizes, and observed that the curvature of the TM domain is enforced by smaller micelle sizes (~ 35 Å). The driving force for the formation of the bend is the solvation of the G₂₉XXXGXXXG₃₇ motif, which is largely exposed to the bulk solvent. The APP-TM is less exposed to the solvent in the larger micelles (~ 45 Å), favoring the straighter conformer. Since different surfactants (DPC and LMPG, respectively) composed the micelles used for the NMR spectroscopy, the APP_{2LLM} structure was probably embedded into larger micelles, producing a much straighter conformation. The exact characteristics of the NMR samples however are difficult to determine. In order to gain insight into their composition, we compared the experimental and back-calculated MD RDCs. The RDC measurements provide long-range orientation restraints and are therefore widely used for NMR structure determination. A genetic algorithm was used to extract an ensemble of structure correctly reproducing the experimental results. A heterogeneous population of straight and bent helices was obtained, suggesting as observed in MD simulations, that the APP-TM backbone can adopt several different conformations. These results are also consistent with the free energy landscape of APP in micelles, where we observed that APP could explore almost freely the conformational space associated to the TM bending.

To give further support to our model, we compared it to the EPR power saturation and DEER experiments used to analyze the structural characteristics of APP-C99 in the POPC liposome (1). EPR power saturation allowed estimating the burial of each residue in the membrane. Based on the MD simulations performed in the POPC membrane, a straight transmembrane helix model better matches the reported EPR power saturation profile (Figure 21b). By enforcing a bend as observed in the APP_{2LP1}

structure, the most buried amino acid shifted by 5 residues. The length of the transmembrane domain was measured by the EPR DEER experiments. Cysteine mutations were performed at positions G700 and L723, in order to covalently bind MTSL spin labels. The MD simulations showed that mutations of G700 significantly affected the dynamical behavior of APP. G700 is the last residue of the loop connecting the juxta-membrane domain to the TM domain. Therefore, perturbing this region directly affects the APP backbone flexibility, causing the juxta-membrane domain to rearrange itself (Figure S10). This increases the solvation of the TM N-terminus, and in turn perturbs the entire TM domain. When the MTSL spin labels are bound to G700C, we observed that the dynamical properties of the TM domain are even more dramatically affected. The curved and straight conformations of the TM helix were observed in the POPC bilayer, with a slight preference for the latter one. The deep burial of the MTSL spin label into the membrane surface forces the APP helix to bend. The MTSL spin labels have shown a preference for hydrophobic cavities (17, 24). In the presence of a membrane, these markers have a tendency to bury themselves into the hydrophobic core of the membrane. The differences observed between the predicted distance distribution and the one extracted from MD simulation could be explained by the fact that the rotamer libraries used for the MTSL prediction algorithms are based on long simulations in solution (17). They are therefore not able to consider the effect of the hydrophobic environment of the membrane. Once buried into the membrane, the spin labels can no longer explore the full conformation space, biasing the expected simulated distance towards shorter distances. The broad distance distribution observed experimentally is caused by the simultaneous presence of the straight and kinked helices, as documented by MD simulations. The Boltzmann distribution of structures derived from the measured free energy landscape, matched the most accurately the experimental results. Consistent with experimental findings, the free energy landscape also confirmed that G708 confers more flexibility to the APP-TM (1). By mutating this residue, the flexibility is reduced, thereby biasing the structure towards a straighter helix. This produces a much sharper DEER distance distribution.

In conclusion, APP appears to be a highly dynamic structure that can easily adapt to changes in the lipid environment. In this context, the atomistic interpretation of experimental data accessible by enhanced sampling molecular dynamics simulations was fundamental in discovering that APP, while having preferential conformers under different conditions, is better described by a dynamic ensemble of conformations that explore the intrinsic flexibility encoded in the G708-G709 motif. Alterations in the membrane composition, especially the concentration of cholesterol, have been linked to AD. It is hypothesized that they affect the formation and insertion of the plaques (25-27). Since a change in the concentration of cholesterol would also affect the flexibility and thickness of the membrane (28), the structure of the APP-TM domain would probably also be affected. In addition, the mutations causing familial AD might affect the flexibility of the TM domain and therefore alter the cleavage profile of the APP. The low-resolution cryo-EM map of γ -secretase suggests a possible arrangement of the α -helical binding site, which might favor a curved structure of the APP substrate (29, 30). Therefore, the curved APP_{2LP1} structure observed in micelles could be better accommodated and bind to the γ -secretase active site (1). The observed flexibility of APP could play two key roles in γ -secretase-mediated

proteolysis. First, the induced shift would be essential for exposing the consecutive cleavage sites to the γ -secretase. Second, the bending could recruit, as in micelles, water molecules required for catalytic cleavage at the hydrophobic core of the membrane.

4.1.4 Material and Methods

Molecular dynamics (MD) simulations were used to characterize the structure of APP in micelles and membrane bilayers, under the effect of different experimental conditions. The NMR structure (PDB code 2LP1, (1)) was inserted and equilibrated in 35 Å and 45 Å lyso myristoyl phosphatidyl glycerol (LMPG) micelles and in a 60 x 60 Å² palmitoyl oleoyl phosphatidyl choline (POPC) patch (31). To test the effect of the spin labels used for DEER spectroscopy, we performed cysteine or MTSL bound to cysteine (CM) point mutations on the APP. In total, seven APP mutants (5 single mutations (G700C, G700CM, G700C, G700CM and G700W) and 2 double mutations (G700C/L723C and G700CM/L723CM) were inserted into a 60 x 60 Å² POPE membrane. All systems were solvated in a 15 Å padding water box, neutralized by the addition of NaCl at a concentration of 150 mM (Table S1).

All simulations were performed using NAMD 2.8 (32) engine, with the CHARMM27 force field (33), including CMAP corrections for the protein and CHARMM36 for the POPC membrane (34). The parameters of the MTSL spin label were extracted from the work of Sezer *et al.* (9). TIP3P water (35) parameterization was used to describe the water molecules. The periodic electrostatic interactions were computed using particle-mesh Ewald (PME) summation with a grid spacing smaller than 1 Å. Constant temperature (300 or 318 K) was imposed by using Langevin dynamics (36) with a damping coefficient of 1.0 ps. Constant pressure of 1 atm was maintained with Langevin piston dynamics (37), 200 fs decay period and 50 fs time constant. All systems were equilibrated for 20 ns at 300 K. Free molecular dynamics were performed up to hundreds of ns timescale with a 2 fs integration time step using the RATTLE algorithm applied to all bonds. The list of simulations is reported in Table S1 for a cumulative simulation time of 2 μ s.

Metadynamics is used to enhance the sampling of the conformational space and to retrieve the associated free energy landscape. Gaussians are added to the energy surface and force the system to escape from local minima. This technique requires *a priori* knowledge of the degrees of freedom relevant to a conformational change. In the current systems, the observed conformational change was described by means of two collective variables representing the dihedral angles ϕ and ψ , which connect the nitrogen atoms of G708 to G709 and G709 to V710. These angles allow the sampling of the different kinked conformations observed for the TM domain in the NMR structures. A set of three metadynamics simulations (~30 ns each, Table S1) was performed using the collective variable module of NAMD. Gaussians of width 0.3 and weight 0.01 were inserted every 300 fs. The conformational space was sampled from -75 to 10 degrees for ϕ and ψ (Figure 19d, Figure 20d, Figure 21e).

The residual dipole couplings (RDCs) were back calculated based on the orientation of the backbone N–H bond vector. A singular value decomposition (SVD) was used

to compute the alignment tensor, which would best map the experimental RDCs. First, the N–H vector was extracted from the APP_{2LP1} ensemble and then back computed the RDCs. A set of structures was then extracted from the MD trajectories performed for the two micelle systems. The structures that would best fit the experimental RDCs were selected with a genetic algorithm. We finally clustered them using the Jarvis-Patrick clustering algorithm (38) (Figure S8).

Three different EPR DEER prediction algorithms were considered in order to estimate the distance distribution profile for the tagged APP-TM domain, namely MMM (21), mtsslWizzard (22), and Pronox (23). The DEER signal was computed for the straight structure equilibrated in a POPC bilayer and the curved NMR structure. A modified version of MMM was used to estimate the distance distribution based on the MTSL coordinates extracted from the MD and metadynamics simulations (Figure 21d and Figure S12).

4.1.5 Supporting Material

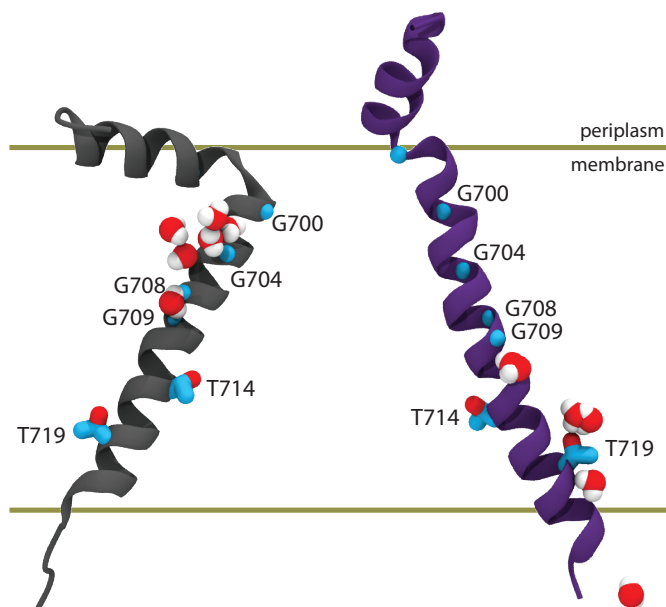


Figure S7: Solvation of the APP-TM. The water molecules access the TM core via the N-terminus (left) or C-terminus (right). The membrane surface is indicated with gold lines.

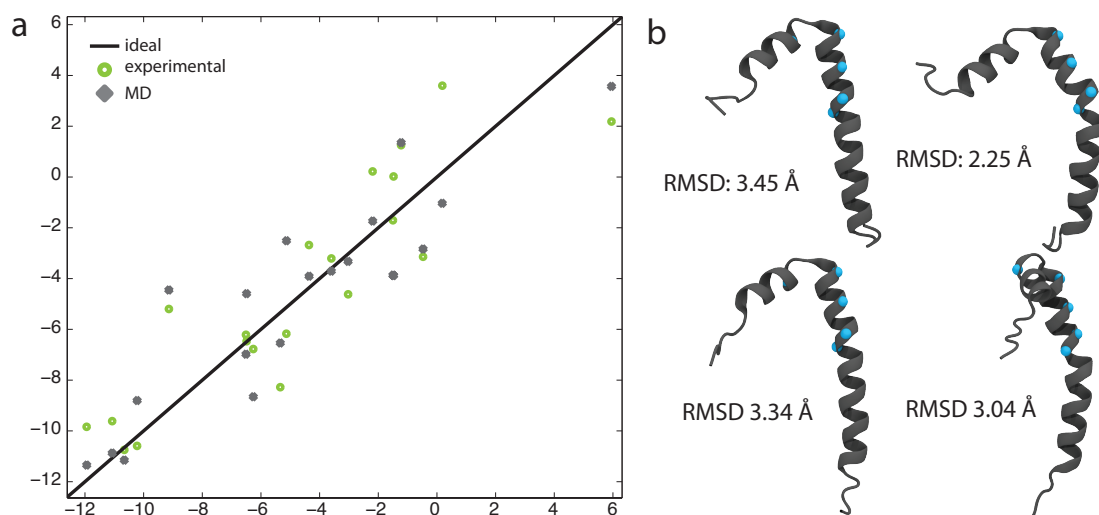


Figure S8: Residual dipole coupling analysis. (a) The correlation of residual dipole coupling (RDC) was back calculated from the NMR ensemble of structures (green) and MD simulations performed in both micelles; M45 and M35 (grey), and (b) The four major conformers are reported with their RMSD with respect to the APP2LP1 NMR structures.

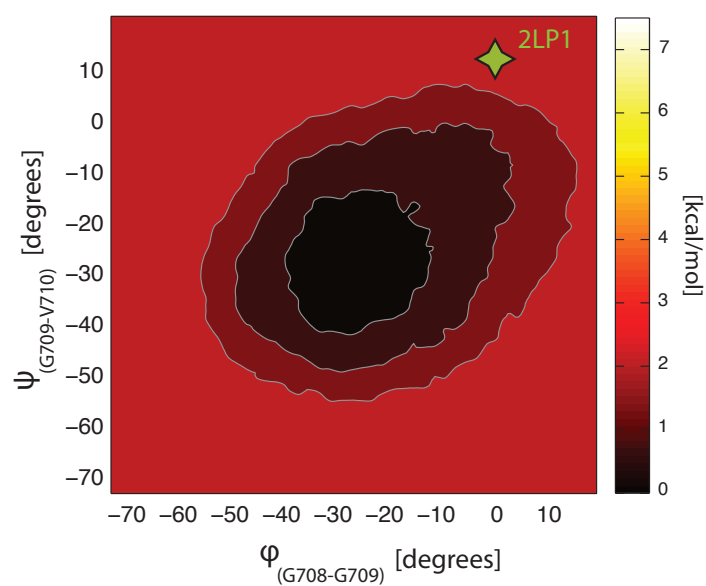


Figure S9: APP Free energy landscape. The free energy landscape associated to dihedral angles ϕ and ψ as calculated by metadynamics MD on the M35 system. The green asterisk indicates the average NMR conformation relative to the 2LP1 ensembles.

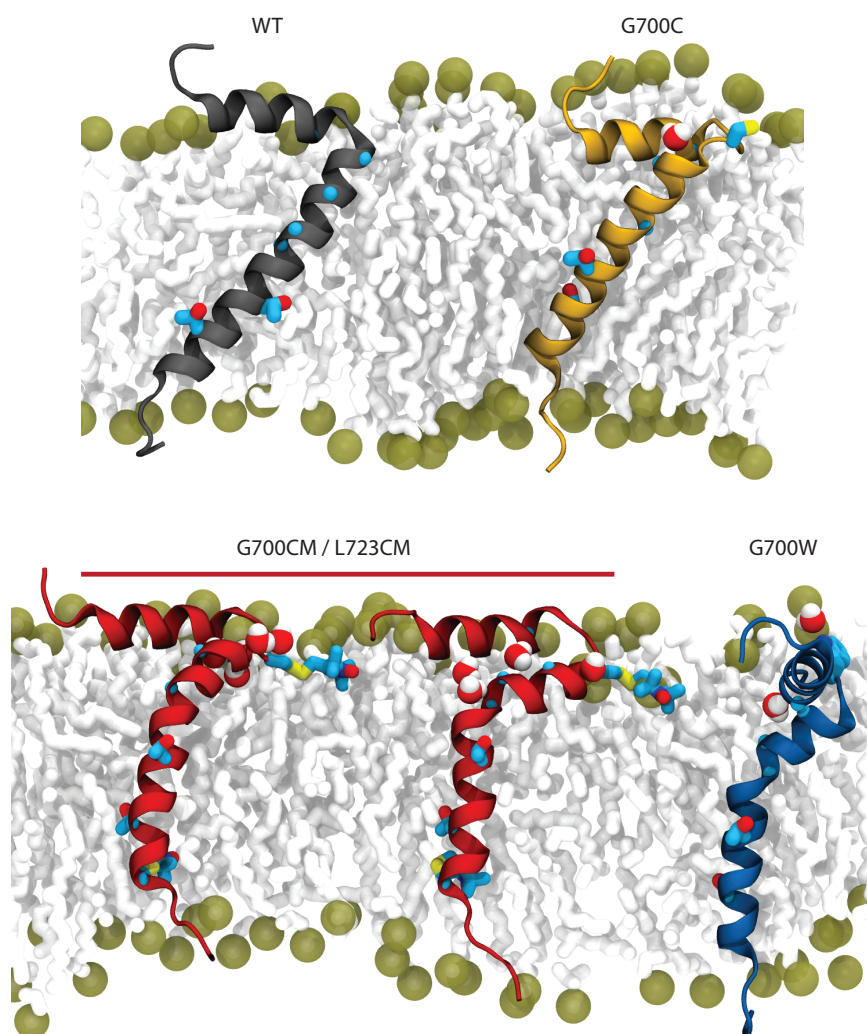


Figure S10: Structural model of APP mutants and spin labeled. The wild type and three relevant mutants, namely G700C, [G700CM/L723CM] and G700W, were inserted in a POPC membrane bilayer and simulated with MD.

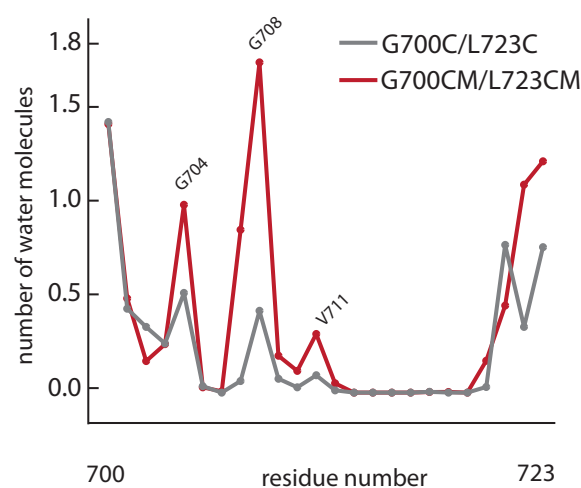


Figure S11: Solvation of APP-TM. The relative level of solvation of APP-TM for two G700 mutations.

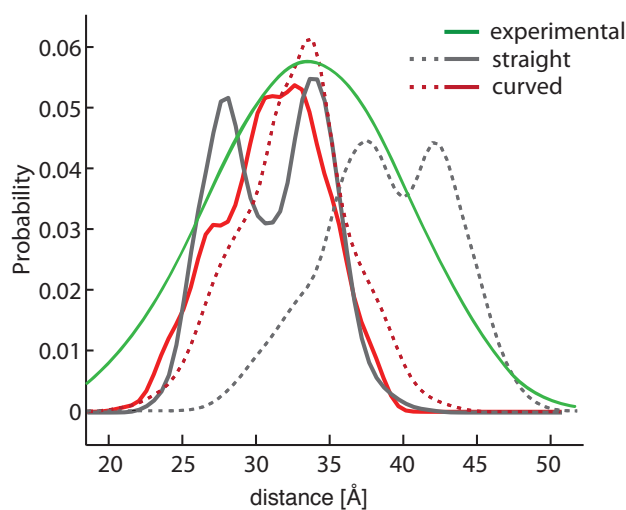


Figure S12: DEER distance distribution. The experimental distance distribution (1) (green) is compared to the predicted distance distribution using MMM (dashed lines) and the distribution measured for MD with MTSL tags (solid lines).

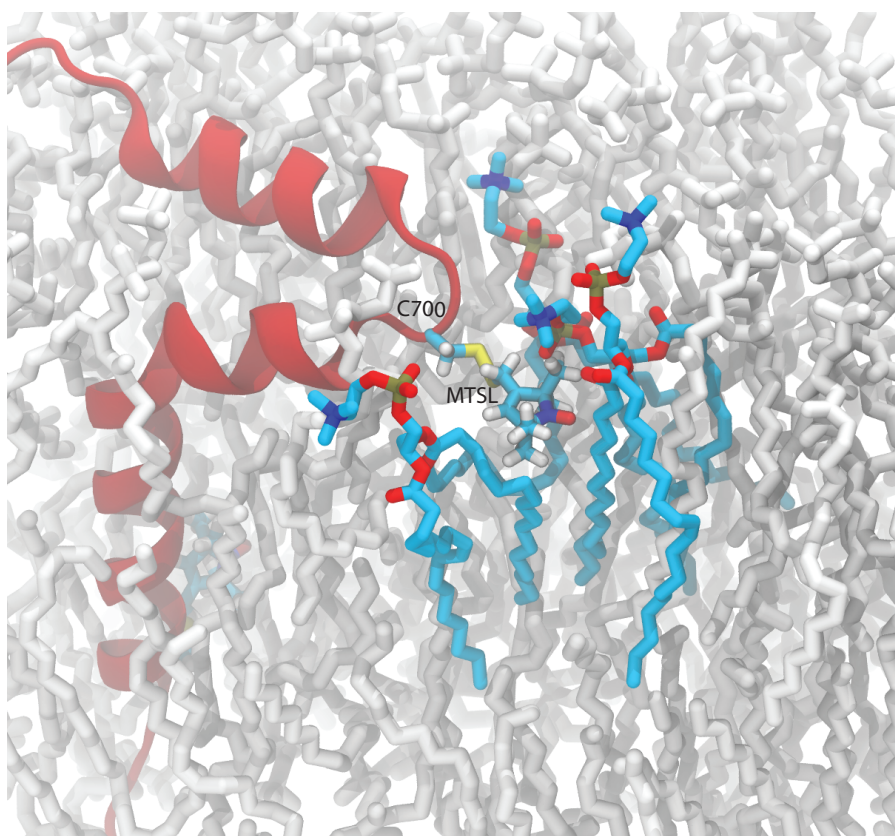


Figure S13: Spin labeled transmembrane domain. Lipids interacting with the MTSL spin label are highlighted in blue.

Table S1. List of performed simulations

system name	lipid type	# of atoms	free MD	metadynamics
WT 2LP1	POPC	40,000	200 ns	100 ns
WT 2LLM	POPC	40,000	200 ns	-
M35	LMPG	65,000	150 ns	30 ns
M35 @ 318K	LMPG	65,000	100 ns	-
M45	LMPG	65,000	150 ns	30 ns
M45 @ 318K	LMPG	65,000	100 ns	-
G700C	POPC	40,000	100 ns	-
L723C	POPC	40,000	100 ns	-
G700C/L723C	POPC	40,000	120 ns	-
G700CM	POPC	40,000	180 ns	30 ns
L723CM	POPC	40,000	100 ns	-
G700CM/L723CM	POPC	40,000	300 ns	2 x 60 ns
G700W	POPC	40,000	130 ns	-

Table S2. DEER distance prediction

system	MMM [Å]	MTSSLWizard [Å]	Pronox [Å]
straight helix	38.9 ± 4.4	40.61	39.9 ± 4.3
bent helix	31.3 ± 3.7	25.61	30.1 ± 4.3

4.1.6 References

1. Selkoe DJ (2001) Alzheimer's disease: genes, proteins, and therapy. *Physiol Rev* 81: 741-766.
2. Barnham KJ, McKinstry WJ, Multhaup G, Galatis D, Morton CJ, et al. (2003) Structure of the Alzheimer's disease amyloid precursor protein copper binding domain. A regulator of neuronal copper homeostasis. *J Biol Chem* 278: 17401-17407.
3. Beel AJ, Mobley CK, Kim HJ, Tian F, Hadziselimovic A, et al. (2008) Structural studies of the transmembrane C-terminal domain of the amyloid precursor protein (APP): does APP function as a cholesterol sensor? *Biochemistry* 47: 9428-9446.
4. Miyashita N, Straub JE, Thirumalai D (2009) Structures of beta-Amyloid Peptide 1-40, 1-42, and 1-55-the 672-726 Fragment of APP-in a Membrane Environment with Implications for Interactions with gamma-Secretase. *Journal of the American Chemical Society* 131: 17843-17852.
5. Russ WP, Engelman DM (2000) The GxxxG motif: a framework for transmembrane helix-helix association. *J Mol Biol* 296: 911-919.
6. Munter LM, Voigt P, Harmeier A, Kaden D, Gottschalk KE, et al. (2007) GxxxG motifs within the amyloid precursor protein transmembrane sequence are critical for the etiology of A beta 42. *Embo Journal* 26: 1702-1712.
7. Sezer D, Freed JH, Roux B (2008) Parametrization, molecular dynamics simulation, and calculation of electron spin resonance spectra of a nitroxide spin label on a polyalanine alpha-helix. *Journal of Physical Chemistry B* 112: 5755-5767.
8. Munter LM, Botev A, Richter L, Hildebrand PW, Althoff V, et al. (2010) Aberrant Amyloid Precursor Protein (APP) Processing in Hereditary Forms of Alzheimer Disease Caused by APP Familial Alzheimer Disease Mutations Can Be Rescued by Mutations in the APP GxxxG Motif. *Journal of Biological Chemistry* 285: 21636-21643.
9. Miyashita N, Sugita Y (2010) Replica-Exchange Molecular Dynamics Simulations of Amyloid Precursor Protein Dimer in Membrane. *Quantum Bio-Informatics lii: From Quantum Information to Bio-Informatics* 26: 361-368.
10. Nadezhdin KD, Bocharova OV, Bocharov EV, Arseniev AS (2011) Structural and dynamic study of the transmembrane domain of the amyloid precursor protein. *Acta Naturae* 3: 69-76.
11. Barrett PJ, Song YL, Van Horn WD, Hustedt EJ, Schafer JM, et al. (2012) The Amyloid Precursor Protein Has a Flexible Transmembrane Domain and Binds Cholesterol. *Science* 336: 1168-1171.
12. Laio A, Parrinello M (2002) Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America* 99: 12562-12566.
13. Cotman C, Blank ML, Moehl A, Snyder F (1969) Lipid composition of synaptic plasma membranes isolated from rat brain by zonal centrifugation. *Biochemistry* 8: 4606-4612.

14. Showalter SA, Bruschweiler R (2007) Quantitative molecular ensemble interpretation of NMR dipolar couplings without restraints. *J Am Chem Soc* 129: 4158-4159.
15. Malmberg NJ, Falke JJ (2005) Use of EPR power saturation to analyze the membrane-docking geometries of peripheral proteins: applications to C2 domains. *Annu Rev Biophys Biomol Struct* 34: 71-90.
16. Frazier AA, Wisner MA, Malmberg NJ, Victor KG, Fanucci GE, et al. (2002) Membrane orientation and position of the C2 domain from cPLA2 by site-directed spin labeling. *Biochemistry* 41: 6282-6292.
17. Jeschke G (2012) DEER Distance Measurements on Proteins. *Annual Review of Physical Chemistry*, Vol 63 63: 419-446.
18. Jeschke G, Polyhach Y (2007) Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance. *Physical Chemistry Chemical Physics* 9: 1895-1910.
19. Pannier M, Veit S, Godt A, Jeschke G, Spiess HW (2000) Dead-time free measurement of dipole-dipole interactions between electron spins. *Journal of Magnetic Resonance* 142: 331-340.
20. Martin RE, Pannier M, Diederich F, Gramlich V, Hubrich M, et al. (1998) Determination of end-to-end distances in a series of TEMPO diradicals of up to 2.8 nm length with a new four-pulse double electron electron resonance experiment. *Angewandte Chemie-International Edition* 37: 2834-2837.
21. Jeschke G, Polyhach Y, Bordignon E (2010) Multiscale Modeling of Macromolecules – MMM. ETH Zurich www.epr.ethz.ch/software/index.
22. Hagelueken G, Ward R, Naismith JH, Schiemann O (2012) MtsslWizard: In Silico Spin-Labeling and Generation of Distance Distributions in PyMOL. *Appl Magn Reson* 42: 377-391.
23. Hatmal MM, Li Y, Hegde BG, Hegde PB, Jao CC, et al. (2012) Computer modeling of nitroxide spin labels on proteins. *Biopolymers* 97: 35-44.
24. Lillington JED, Lovett JE, Johnson S, Roversi P, Timmel CR, et al. (2011) Shigella flexneri Spa15 Crystal Structure Verified in Solution by Double Electron Electron Resonance. *Journal of Molecular Biology* 405: 427-435.
25. Pettegrew JW, Panchalingam K, Hamilton RL, McClure RJ (2001) Brain membrane phospholipid alterations in Alzheimer's disease. *Neurochem Res* 26: 771-782.
26. Ji SR, Wu Y, Sui SF (2002) Cholesterol is an important factor affecting the membrane insertion of beta-amyloid peptide (A beta 1-40), which may potentially inhibit the fibril formation. *J Biol Chem* 277: 6273-6279.
27. Bokvist M, Lindstrom F, Watts A, Grobner G (2004) Two types of Alzheimer's beta-amyloid (1-40) peptide membrane interactions: aggregation preventing transmembrane anchoring versus accelerated surface fibril formation. *J Mol Biol* 335: 1039-1049.
28. Buchet R, Pikula S (2000) Alzheimer's disease: its origin at the membrane, evidence and questions. *Acta Biochim Pol* 47: 725-733.
29. Osenkowski P, Li H, Ye WJ, Li DY, Aeschbach L, et al. (2009) Cryoelectron Microscopy Structure of Purified Y-Secretase at 12 angstrom Resolution. *Journal of Molecular Biology* 385: 642-652.

30. Renzi F, Zhang XL, Rice WJ, Torres-Arancivia C, Gomez-Llorente Y, et al. (2011) Structure of gamma-Secretase and Its Trimeric Pre-activation Intermediate by Single-particle Electron Microscopy. *Journal of Biological Chemistry* 286: 21440-21449.
31. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14: 33-38, 27-38.
32. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 26: 1781-1802.
33. Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr., Nilsson L, Petrella RJ, et al. (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30: 1545-1614.
34. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, et al. (2010) Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *Journal of Physical Chemistry B* 114: 7830-7843.
35. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 79: 926.
36. Brunger A, Brooks CL (1984) Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chemical Physics Letters* 105: 495-500.
37. Feller SE, Zhang Y, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation: the Langevin piston method. *The Journal of Chemical Physics* 103: 4613.
38. Jarvis RA, Patrick EA (1973) Clustering Using a Similarity Measure Based on Shared Near Neighbors. *Computers, IEEE Transactions on C-22*: 1025-1034.

4.2 Alzheimer's disease mutations in the Amyloid precursor protein

Revised excerpt from the following paper:

“Early-onset Alzheimer's disease mutations in APP, but not γ -secretase modulators, affect epsilon-cleavage-dependent AICD production”

Dimitrov M., Alattia J., Lemmin T., Rajwinder L., Fligier A., Houacine J., Hussain I., Radtke F., Dal Peraro M., Behr D., and Fraering P. C., submitted

4.2.1 Introduction

The main pathological hallmark of Alzheimer's disease (AD) is the formation of toxic amyloid fibers that progressively accumulate into plaques. Two forms of AD exist: the sporadic form and the genetic form. Several inherited mutations causing the genetic form, also called Familial Alzheimer's disease (FAD), are found in the amyloid precursor protein (APP). It has been shown that pathological amino acid substitutions in the APP and chemical γ -secretase modulators (GSMs) affect the processing of the APP by the γ -secretase complex. The mutations in the transmembrane domain of the APP causing aggressive early-onset familial Alzheimer's disease affect both the γ - and ϵ -cleavage sites, by increasing the A β 42/40 ratio. The ϵ -cleavage site is defined as the first step of the processing of APP and the γ -site as the last one.

In this study, three different FAD mutants were considered (T714I, V715A and V717F). These mutations were chosen because they either affect both the γ - and ϵ -cleavage sites (T714I, V715A), or only the ϵ -site (V717F). The mutations T714I and V715A increased the A β 42/40 and A β 38/40 ratios. In contrast, for V717F, these profiles were comparable to those obtained with wild type (WT) APP. Two ϵ -cleavage sites were observed in the WT: Leu49 and Val50. For all FAD mutants, only the Leu49 site is present. To rationalize these findings, we used molecular dynamics simulations of the APP-TM segment embedded in the lipid bilayer, and proposed that FAD-APP mutants and possibly GSMs affect the γ -cleavage site by changing the flexibility, tilt angle and position of the substrate in the membrane. The FAD mutants also altered the solvation at the ϵ -cleavage site, which can explain the different product lines generated after the first step of the processing of APP.

4.2.2 Results

We used molecular dynamics (MD) simulations to gain insights into the structural and dynamic properties of the transmembrane domain (TMD) of wild-type APP and three FAD-mutants (T714I, V715A and V717F) in the membrane environment. The TM region of the APP-C99 NMR structure (PDB code 2LP1 [1]) was used and inserted into a phosphatidylcholine (POPC) lipid bilayer (Figure 22a). MD simulations were carried out under physiological-like conditions (i.e., pH ~7, 1 atm, 300 K), mimicking as close as possible the conditions of the γ -secretase enzymatic assay used to study

the cleavage. For all systems, the secondary structure of the APP fluctuated only marginally. Consistent with the NMR observations, the N-terminus amphiphilic helix (Q686–N698) lies on the membrane surface, whereas a lysine belt (K687, K699, K724–726) anchors the TMD and N-terminus helices, respectively, in and onto the membrane (Figure 22a). During simulation, the helices tilted with respect to the membrane surface. We found that the V717F mutation affected the orientation of the TMD the most, tilting it by 45 ± 5 degrees, whereas the wild type and T714I and V715A mutants remained more vertical (36 ± 5 , 38 ± 3 and 36 ± 3 degrees, respectively) (Figure 22a and Figure 23a). These mutations further shifted the TMD perpendicularly to the membrane surface. When measuring the position of the A β 42 cleavage site (i.e., I712 – A713 peptide bond), the T714I and V715A mutants shifted closer (2.9 ± 1.2 Å and 2.3 ± 0.9 Å, respectively) to the cytoplasmic interface compared to the WT, and V717F moved closer (-0.2 ± 1.3 Å) to the extracellular membrane layer (Figure 22b).

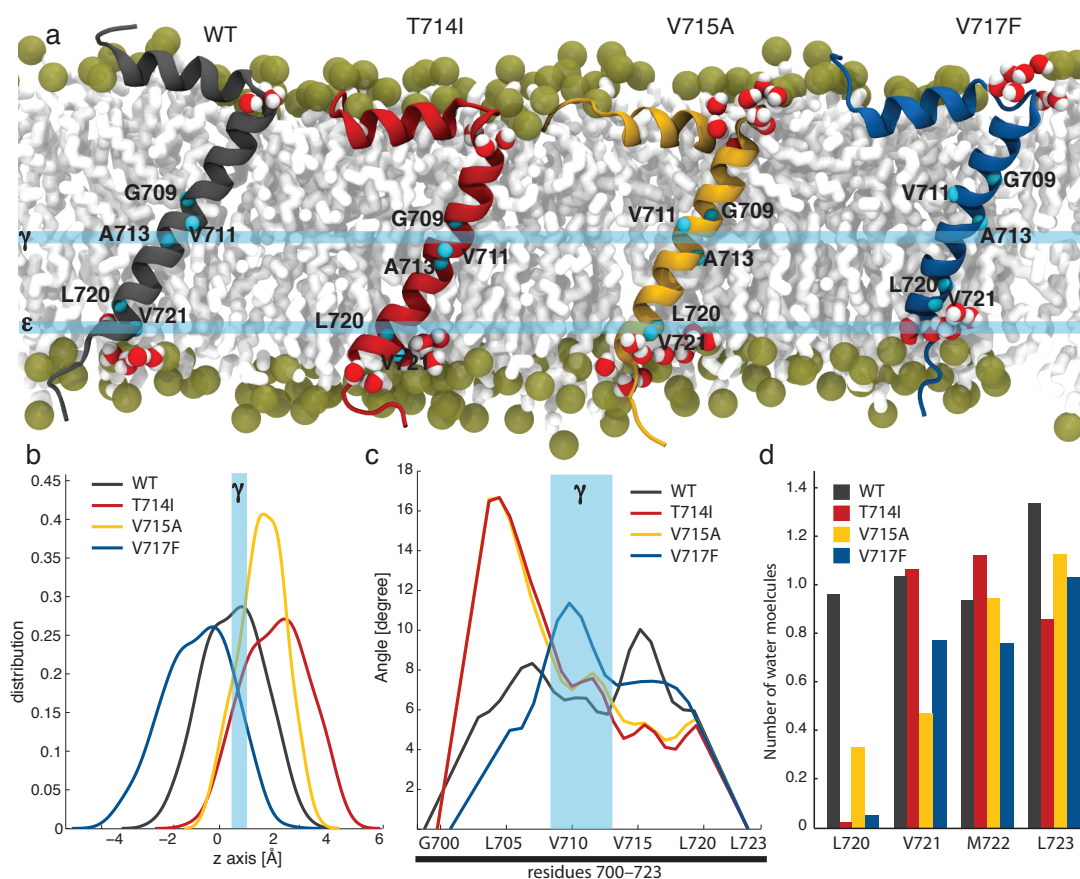
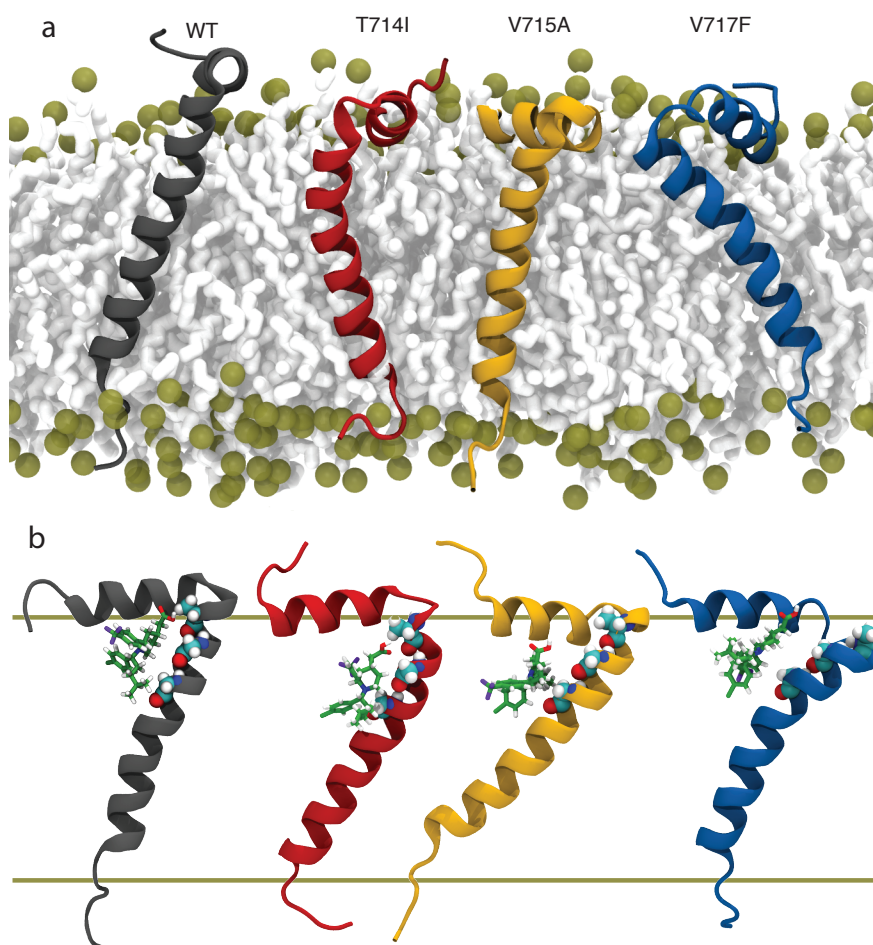


Figure 22: Structural and dynamic models of the TM domain in WT and FAD-causing APP mutants. (a) The NMR structure of WT APP (PDB code: 2LP1 [1]) and three FAD-causing mutants, namely T714I, V715A and V717F, were inserted in a POPC membrane bilayer and simulated with molecular dynamics. The γ -sites G709, V711 and A713 and ϵ -sites L720 and V721 positions are indicated by blue spheres, (b) The spatial distribution of the A β 42 cleavage site is reported for the four systems. The distribution is calculated along the axis orthogonal to the membrane bilayer surface and the reference putative position of the γ -

site in wild-type conditions is highlighted with the light blue area, (c) The local bending of the APP helix is computed for the TM domain (with a standard error of 2.8 degrees) and (d) The coordination number of water molecules around the relevant residues at the ϵ -site is reported for the WT and mutants.

The combined effect of shifting and tilting perturbations of the APP TMD due to FAD-mutations probably affects the presentation of the γ -cleavage site at position 42 to the enzyme. This correlates quantitatively with the changes in the A β 42/ A β 40 ratio. Moreover, the flexibility of the transmembrane domain, that was estimated based on the local bending of the helix [2], was much higher for the T714I and V715A mutants, specifically at the G₂₉XXXGXXXG₃₇ region (Figure 22c). Molecular docking and simulations of the GSM-1 compound suggested that the drug-binding pocket is conserved for the wild-type and mutant APPs (Figure 23b). In the drug-binding pocket, L701, G704 and G708 had the most direct interactions with GSM-1. Thus, the effect of GSM-1 on T714I and V715A FAD mutants might cause a reduction of their flexibility close to the γ -site. For all three mutants, the hydration of the ϵ -site at residue L720 significantly decreased compared to the WT (Figure 22d). In particular, for APP-WT, one water molecule is on the average within 3 Å of L720 backbone, the water coordination number drops to less than 0.5 Å for V715A, and to nearly zero for T714I and V717F. This effect is related to the different water permeation properties of the APP C-terminus lysine belt combined with the perturbed helical orientation produced by the mutants. In summary, our data support a model in which structural changes, in particular, the position of the γ -cleavage site relative to the membrane surface potentially can modulate A β production, while the hydration of L720 can affect the ϵ -cleavage site and the production of the two amyloid precursor protein intracellular domains potentially implicated in signalling pathways and in the pathogenesis of AD.

Figure 23: Molecular docking and MD simulations of GSM-1 bound to APP-C99. (a) Different orientations in the membrane of APP WT and mutants TMD, (b) GSM-1 binding to WT, T714I, V715A and V717F APP obtained using SWISSDock and BSP-SLIM revealed that the compound binds to the G29XXXGXXXG37 motif (in ball representation) with a conserved and stable conformation. The wild-type system is equilibrated using MD simulation and confirmed a stable pose of GSM-1 compound in dynamics.



4.2.3 Discussion

We propose that the observed effects on A β production can be triggered by combined changes in the flexibility, tilting angle, and positioning of both the γ - and ϵ -sites in the membrane, along with the hydration levels at the ϵ -sites. Such structural changes might affect the γ -secretase cleavage specificity by potentially modifying the presentation of the substrates to the enzymatic complex. Consistent with this model, the high A β 38/40 ratio observed in mutants T714I and V715A can be explained by the vertical shift of the helix in the membrane bilayer, with the consequent presentation of position 38 instead of 40 to the γ -secretase cleavage site. The observed vertical shift in mutants (~ 2.5 Å) is in fact associated with a tilting of half of a helix turn (i.e., displacement of 2 amino acids).

The increased water accessibility to L49 may favor cleavage at this position, or inhibit cleavage at V50 by preventing the interaction of APP-C99 at position 50 with the hydrophobic transmembrane domain 1 of the γ -secretase catalytic.

To better understand the effect of GSM binding, the GSM-1 compound was docked on the ensemble of wild-type and FAD-mutant (T714I, V715A and V717F) conformations were generated by MD simulations. GSM-1 strongly binds to the

previously identified G₂₉XXXGXXXG₃₇ motif (Figure 23b), further supporting our observation that amino acids close to the γ -sites of APP are not required for GSMs activity, again underlying the therapeutic potential of these compounds [3].

In conclusion, molecular simulations suggest that FAD mutations modulate the flexibility of the APP transmembrane domain. The presentation of its γ -site and the solvation of the ε -site are thus modified.

4.2.4 Material and Methods

Molecular dynamics simulations and molecular docking

We used molecular dynamics (MD) simulations to characterize the TM domain of WT and APP with FAD mutations. In a first step, the NMR structure (PDB code 2LP1^[1]) was inserted and equilibrated in a 60 x 60 Å² Palmitoyl Oleoyl Phosphatidyl Choline (POPC) patch^[4] to relax the structure in a phospholipid bilayer. The system was solvated in a 60 x 60 x 100 Å³ water box, neutralized through the addition of NaCl at a concentration of 150 mM. After 30 ns, the equilibrated structure was used as initial seed to build the other three mutants (T714I, V715A and V717F) and WT-ligand complex. GSM-1 ligand was parameterized using CGenFF and ParamChem 0.9.6^[5]. All MD simulations were performed using NAMD 2.8 engine, with the CHARMM27 force field, including CMAP corrections for the protein and CHARMM36 for the POPC membrane [6,7]. TIP3P water parameterization was used to describe the water molecules. The periodic electrostatic interactions were computed using the particle-mesh Ewald (PME) summation with a grid spacing smaller than 1 Å. All systems were equilibrated for 30 ns at 300 K. Free molecular dynamics were performed up to 100 ns with a 2 fs integration time step using the RATTLE algorithm applied to all bonds. Constant temperature (300 K) was imposed by using Langevin dynamics, with a damping coefficient of 1.0 ps. A constant pressure of 1 atm was maintained with Langevin piston dynamics, having a 200 fs decay period and a 50 fs time constant.

GSM-1 was docked on the WT and APP mutants using SWISSDock [8,9] and BSP-SLIM [10]; independently both approaches produced similar configurations. Since these algorithms do not account for the presence of the membrane, we discarded all unphysical docking poses, e.g. the GSM-1 hydroxyl was buried in the hydrophobic membrane core. The conformation with the lowest energy was then selected for each system.

4.2.5 References

1. Barrett PJ, Song Y, Van Horn WD, Hustedt EJ, Schafer JM, et al. (2012) The amyloid precursor protein has a flexible transmembrane domain and binds cholesterol. *Science* 336: 1168-1171.
2. Dahl AC, Chavent M, Sansom MS (2012) Bendix: intuitive helix geometry analysis and abstraction. *Bioinformatics* 28: 2193-2194.
3. Kukar TL, Ladd TB, Bann MA, Fraering PC, Narlawar R, et al. (2008) Substrate-targeting gamma-secretase modulators. *Nature* 453: 925-929.
4. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14: 33-38, 27-38.
5. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, et al. (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* 31: 671-690.
6. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26: 1781-1802.
7. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, et al. (2010) Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J Phys Chem B* 114: 7830-7843.
8. Grosdidier A, Zoete V, Michielin O (2011) SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res* 39: W270-277.
9. Grosdidier A, Zoete V, Michielin O (2011) Fast docking using the CHARMM force field with EADock DSS. *J Comput Chem*.
10. Lee HS, Zhang Y (2012) BSP-SLIM: a blind low-resolution ligand-protein docking approach using predicted protein structures. *Proteins* 80: 93-110.

Conclusions and Perspectives

The biological membrane forms a nearly impermeable barrier, defining the boundaries of a living organism. Its lipid composition closely affects the stability and mechanism of membrane-embedded proteins, causing a large disparity between the understanding of soluble and membrane-proteins. To overcome these issues, detergent micelles are often used during protein purification, crystallization and functional studies. However, these experimental setup can strongly influence the structural and dynamical determinants of the protein. Moreover, the conformational changes essential for the function, the stability or the binding of the protein are difficult to deduce from wet experiments. Molecular modeling is therefore an essential tool for gaining insight into the structure and function of membrane-proteins. The computational power available today allows simulating large and complex systems, which are closely mimicking realistic models of protein-membrane complexes.

This thesis project dealt with the following four important problems related to the study of membrane-embedded proteins at the atomistic level and proposed various solutions:

1. The determination of the high-resolution structure of transmembrane proteins is difficult. A method for the assembly of transmembrane proteins, combining molecular modeling with structural and experimental studies, was proposed. By integrating secondary predictions and cross-linking data to molecular dynamics simulations, a model of the transmembrane domain of PhoQ was assembled. This model was fully consistent with the available adjoining structures and a signal transduction mechanism could be formulated.
2. Fine interactions between the membrane-proteins and their environment are difficult to address experimentally. Molecular dynamics simulations were used to study the interactions between the membrane and the embedded proteins at the atomistic scale. Modeling of a large diversity of lipids is therefore required. Two aspects of this problem were dealt with. First, the parameterization of cardiolipin, a unique dimeric phospholipid, supplemented the available head group models in the CHARMM and AMBER force fields. This anionic lipid is essential for the investigation of bacterial and mitochondrial membranes. The assembly of realistic membranes can be accelerated by facilitating the creation of an *ad hoc* mixture of lipids. Based on the SMILE sequence of the acyl chain, a software automatically generating the topology and template of the required phospholipid was implemented. It is currently being ported to a web server, allowing easy access and the establishment of a lipid database.

3. Adapting characterization methods for solution proteins to membrane proteins is not always straightforward. The lipid bilayer could strongly affect the expected measurements. In this thesis project, molecular dynamics simulations were shown to be an important complement for the rationalization and refinement of such experimental data. The effect of the membrane on different experimental setups was investigated at the atomistic scale. The dynamical behavior of the transmembrane domain was also deduced by combining MD simulations with the analysis of data such as cross-linking, NMR or EPR.

4. The slow diffusion and conformational properties of the embedded proteins require long simulation time. Classical molecular dynamics at present are not able to sample the entire conformational space within a reasonable timeframe. Enhanced sampling techniques are necessary to more accurately interpret certain experimental data. Metadynamics simulations were used to characterize the dynamical properties of transmembrane proteins and were essential to accurately explain the corresponding *in vitro* measurements.

The organization of the biological membrane at the nanometer scale is difficult to study experimentally. At present, various techniques permit investigations at the molecular scale (e.g. spectroscopy) and at the micrometer scale (e.g. microscopy). Molecular modeling has proven to be a valuable tool for bridging these two experimental techniques. The work carried out in this thesis project effectively integrated MD simulations and experimental data, thus contributing to the understanding of membrane-proteins properties. However, membrane-proteins often assemble in large complexes and require simulation times that are still beyond the reach of today's computational power. Novel strategies, such as enhanced sampling (applied in this thesis) and coarse graining, would be required to study these complexes. Moreover, combining molecular dynamics simulations with other computational techniques can further improve the sampling. For example, molecular modeling based on multistate Bayesian scoring functions has recently proven to be an efficient method. It allows combining data available from different experimental procedures, thus that molecular dynamics simulations can be integrated to these Bayesian strategies in three ways: (i) The dynamic behavior of high-resolution structures of subunits composing a molecular assembly can be determined using MD and then integrated as extra restraints to the Bayesian modeling, (ii) The properties of experimental probes can be significantly altered due to their environment (e.g., membrane). By simulating an atomistic model representing the experimental set-up, Bayesian models of the experimental data can be improved, and (iii) Initial modeling of assemblies is usually performed using a residue-based coarse-grained representation. Biased molecular dynamics can be used to reconstruct the all-atom representation of these models. For example, Molecular Dynamics Flexible Fitting (MDFF) has proven to be an efficient method for fitting an atomic model into a low-resolution density map. In a similar way, the coarse-grained models produced can be used to guide the assembly of a near-atomic model during MD simulations.

At present, computational methods generate structural models mostly using literature data to guide computation. However, modeling the structure and dynamics of macromolecular assemblies allows gaining valuable insight into important biological functions, interactions and conformational changes. This feedback could be used to

stimulate new experiments or, better, to be directly incorporated when designing experiments aimed at unveiling the biological function of molecular complexes.

Annexe

Additional publications and studies related to this thesis:

- *Early-onset Alzheimer's disease mutations in APP, but not γ -secretase modulators, affect epsilon-cleavage-dependent AICD production,*
Dimitrov M., Alattia J., Lemmin T., Rajwinder L., Fligier A., Houacine J., Hussain I., Radtke F., Dal Peraro M., Beher D., and Fraering P. C.
- *Differential effects of frameshift mutations in the exon13 hotspot leading to Hyaline Fibromatosis Syndrome,*
Yan S., Lemmin T., Salvi S., Lausch E., Superti-Furga A., Dal Peraro M., and van der Goot F. G.
- *Palmitoylated calnexin is a key component of the ribosome-translocon complex,* EMBO Journal, 2012
Lakkaraju A.K.K., Abrami L., Lemmin T., Blaskovic S., Kunz B., Kihara A., Dal Peraro M., and van der Goot F. G.

Early-onset Alzheimer's disease mutations in APP, but not γ -secretase modulators, affect epsilon-cleavage-dependent AICD production.

Mitko Dimitrov¹, Jean-René Alattia¹, Thomas Lemmin², Lehal Rajwinder³, Andrzej Fligier¹, Jemila Houacine¹, Ishrut Hussain⁴, Freddy Radtke³, Matteo Dal Peraro², Dirk Beher⁴, and Patrick C. Fraering^{1*}

¹ Brain Mind Institute and School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH1015 Lausanne, Switzerland

²Institute of Bioengineering and School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

³Swiss Institute for Experimental Cancer Research and School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH1015 Lausanne, Switzerland

⁴Global Research and Early Development, Merck Serono SA, 9 Chemin des Mines, 1202 Geneva, Switzerland

* Correspondence should be addressed to Patrick C. Fraering, Brain Mind Institute and School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH1015 Lausanne, Switzerland. E-mail: patrick.fraering@epfl.ch

ABSTRACT

Pathological amino acid substitutions in the amyloid precursor protein (APP) and chemical γ -secretase modulators (GSMs) affect the processing of APP by the γ -secretase complex, and the production of the amyloid-beta peptide A β 42, the accumulation of which is considered causative of Alzheimer's disease. We demonstrate here that mutations in the transmembrane domain of APP causing aggressive early-onset familial Alzheimer's disease (FAD) affect both γ - and ϵ -cleavage sites, by raising the A β 42/40 ratio and inhibiting the production of AICD50-99, one of the two physiological APP intracellular domains (ICDs). This is in sharp contrast to GSMs, which shift A β 42 production towards the shorter A β 38, but unequivocally spare the ϵ -site and APP- and Notch-ICDs production. Molecular simulations suggest that FAD mutations modulate the flexibility of the APP transmembrane domain and the presentation of its γ -site, modifying at the same time the solvation of the ϵ -site.

INTRODUCTION

The main pathological hallmark of Alzheimer's disease (AD) is the formation of toxic amyloid fibers that progressively accumulate into plaques, which leads to degeneration of the neighboring neuronal cells¹. Further biochemical studies identified the main protein component of plaques to be a short fragment of APP, the 42 amino acid long peptide A β 42². Moreover, dominantly inherited mutations causing Familial Alzheimer's disease (FAD) are found in APP as well as in the PSEN1 and PSEN2^{3,4} genes encoding for the catalytic subunit of the γ -secretase complex, an intramembrane-cleaving protease which cleaves APP in the plasma membrane and generates the A β peptides⁵. Together, these findings clearly demonstrated the implication of APP, γ -secretase and A β in both the sporadic and genetic forms of AD. They further highlighted the γ -secretase-dependent APP processing as an attractive target for drug development.

In addition to APP, γ -secretase cleaves various other substrates regulating biological functions including development and stem cell differentiation. This unusual proteolytic complex indeed plays an important role in cell signaling pathways by processing different type-I transmembrane cell receptors including the Notch receptors, the Insulin receptor, and the Growth hormone receptor⁶. In the brain, it maintains synaptic contacts by regulating cleavage of neuronal adhesion molecules like Cadherins and Neurexins^{7,8}. With the continuously growing number of substrates the need for understanding the exact molecular mechanism of this particular proteolytic processing arises.

γ -Secretase is composed of four subunits, presenilin (PS), Aph-1, Pen-2 and Nicastrin (NCT). The complex contains in total 19 α -helices embedded in the membrane bilayer, which allows the overall structure to adopt different conformations upon binding with the substrate or transition-state analog inhibitors⁹. Moreover, the catalytic aspartate residues (Asp257/385) in PS1 are buried into the lipid bilayer and surrounded by hydrophobic helices and the cryo-electron microscopy density maps suggested the existence of two pores in the globular structure of γ -secretase, which can explain the entrance of water molecules necessary for the hydrolysis of the peptide bond¹⁰. Yet, a rigorous biochemical characterization of γ -secretase complexes with FAD mutations revealed that they all represent a loss of enzymatic function, but they ultimately lead to altered A β production and elevated A β 42/40 ratio^{11,12}. Interestingly, FAD mutations in APP can show substrate cleavage patterns that differ from those of mutations in γ -secretase. The large extracellular domain of APP is first removed by β -secretase in the vicinity of the plasma membrane. Gain-of-function mutations in this β -site cause elevated levels of total Ab¹³. γ -secretase processes the remaining 99-amino-acid APP-C terminal fragment (APP-C99) at ϵ -sites close to inner layer of the membrane, and releases 49- and 50-amino-acid long APP- intracellular domains (AICD50-99 and AICD49-99). After this step, γ -secretase sequentially continues to trim the residual transmembrane stub, finally producing A β 42, A β 40 and A β 38 peptides¹⁴. Mutations surrounding the ϵ -sites either increase the production of A β 42, which is more neurotoxic compared to the shorter and less

fibrillogenic A β 40, or lead to elevated total amounts of the A β product without affecting A β ratios^{12,15}. Amino acid substitutions in the g-site lead to high A β 42/40 ratios and are characterized by the most prominent and earliest phenotypes of FAD¹⁶.

Inhibiting the γ -secretase activity to reduce A β production became an attractive therapeutic strategy to treat AD. However, γ -secretase inhibitors (GSIs) tested in clinical trials not only did not meet their initial clinical endpoints, but also manifested severe off-target effects due to impaired Notch signaling, leading to gastrointestinal bleeding and increased risk of skin cancer¹⁷. Consequently, a concept of selective modulation of the γ -secretase activity has emerged. A subset of non-steroidal anti-inflammatory drugs (NSAIDs) were the first Notch-sparing small molecules shown to selectively lower A β 42 and subsequently raise A β 38, both *in vitro* and *in vivo*^{18,19}. These γ -secretase modulators (GSMs) were initially shown to bind to APP, although some more recent studies demonstrated that they also can target γ -secretase^{20,21}. Although the precise molecular mechanism of action remains poorly understood, a number of these molecules have advanced to clinical trials²².

In this report, we demonstrate that second generation GSMs, including the clinically tested E2012, effectively lower the production of toxic A β 42, without affecting the e-cleavage sites of both APP and Notch substrates. In contrast, we show that the T714I, V715A, V717F and L723P FAD-APP mutations in the transmembrane (TM) region block the production of AICD50-99, the 49-amino-acid long AICD, without affecting the total amount of AICD. Moreover, mutations T714I and V715A elevated the A β 42/40 and A β 38/40 ratios, in contrast to I716V, V717F and L723P for which these profiles were comparable to those obtained with wild type APP. GSMs affected all mutants at their γ -cleavage site, with the strongest effect on the A β 38/40 and A β 42/40 ratios in T714I and V715A, suggesting that these amino acid substitutions are important for the generation of A β 42 and A β 38. In addition, we found that GSMs do not reverse the loss of AICD50-99 production caused by FAD mutations. We rationalize these findings using molecular dynamics simulations of the APP TM segment embedded in the lipid bilayer, and propose that FAD-APP mutants and possibly GSMs affect the γ -cleavage site by changing the flexibility, tilt angle and position of the substrate in the membrane. GSMs fail to recover e-cleavage, most likely because they affect the substrate-enzyme interface at a position which is distant from the ϵ -site and closer to the outer layer of the membrane. Finally, FAD mutants had altered solvation at the e-cleavage site, which can explain the different product lines generated after the first step of the processing of APP.

RESULTS

GSMs affect the g- but not the e-cleavage of APP and Notch.

We first investigated the effect of GSM-1 and E2012, two representative molecules of the two main classes of second generation GSMs: NSAID-based phenylpiperidine-type compounds and non-NSAID bridged aromatics, respectively (Fig. 1a). Although these compounds have different structures, they both modulate γ -secretase activity at low nanomolar concentrations in cells²³. In enzymatic *in vitro* assays using highly purified γ -secretase and recombinant APP-C100-Flag substrate, we first show that concentrations several fold higher than the cellular IC₅₀ concentrations of GSM-1 and E2012 do not affect the overall amounts of A β and AICD, when compared to control levels (Fig. 1b, upper panels). We next investigated in detail the effect of these molecules on all reaction products. For both compounds tested at 10 mM, immunoprecipitation with 4G8 anti-A β antibody, combined with MALDI-TOF spectrometry (IP/MS) revealed a robust increase in A β 38, associated with a decrease in A β 42 (Fig.1b, middle panels). At this concentration GSM-1 showed a stronger effect than E2012 on A β 42/40 and A β 38/40 ratios, while both compounds lowered A β 43. IP/MS analyses of the C-terminal cleavage products revealed an unchanged equimolar production of the two AICDs (Fig. 1b, lower panels). Together, these results clearly demonstrate that GSMs affect the APP γ -cleavage sites, while completely sparing the ϵ -sites. Because of

the clinical relevance of GSMs and the Notch signaling pathway, we further questioned whether GSMs could interfere with the Notch intracellular domain (NICD) production. As shown in Figure 1c, GSMs did not influence total NICD production and did not affect the NICD cleavage profiles. Our *in vitro* findings were further confirmed in cell-based assays. First, intracellular accumulation of endogenous γ -secretase substrates APP-C83 and APP-C99, which is a common characteristic among known γ -secretase substrates in response to γ -secretase inhibition, was assessed in HEK293T cells stably overexpressing APP with the FAD Swedish mutation KM670/671NL. In contrast to the control GSI, 1 and 10 mM E2012 or GSM-1 did not affect the levels of APP-CTFs (Supplementary Fig. 1a). Next, treatments with E2012 and GSM-1 up to 1 mM in a cell-based luciferase assay in which AICD generation correlates with increased luciferase expression and induced luminescence emission, did not result in a significant inhibition of AICD release (Supplementary Fig. 1b). Additionally, the GSMs were tested in human lymphocytic leukemia SUP-T1 cells constitutively producing NICD. In contrast to the GSI XIX which completely abolished NICD production, E2012 and GSM-1 did not affect NICD production at concentrations up to 10 mM (Supplementary Fig. 2a). These findings were confirmed in a cellular Notch-based assay in which a luciferase reporter system is directly placed under the control of NICD²⁴ (Supplementary Fig. 2b), as well as in HEK cells overexpressing NotchDE²⁵ (Supplementary Fig. 2c). In the latter two assays, E2012 and GSM-1 concentrations up to 1 mM did not affect NICD production. At 10 mM, E2012, but not GSM-1, significantly reduced NICD production in DL4-Hela/N1-Hela co-cultures by 55% (**p<0.001) (Supplementary Fig. 2b). Taken together, we conclude that E2012 and GSM-1 do not alter the Notch and APP ϵ -site cleavage at concentrations as high as 10x cellular IC₅₀. At higher concentrations (100x cellular IC₅₀), only the non-NSAID-based compound E2012 can inhibit to some extent intracellular signaling of Notch and APP in cell-based assays.

APP mutations causing familial Alzheimer's disease differentially alter g- and ϵ -sites.

APP mutations causing familial Alzheimer's disease (FAD) can be grouped based on their localization at three different regions: at the b-secretase cleavage site, in close vicinity of the a-secretase cleavage site in the A β region, or in the transmembrane domain (TMD) of APP, between the g- and the ϵ -sites (Fig. 2a, left). These mutations differ not only by the onset of the disease (Fig. 2a, right), but also by the molecular fingerprint of the produced A β peptides. Indeed, the vast majority of mutations in the a- and b-sites lead to overproduction of total A β (mutations in the A β region can also render the A β 42 peptide more fibrinogenic and prone to aggregation) while mutations in the transmembrane region seem to be linked to an elevated A β 42/40 ratio^{12,26}. Because no information exists about how these mutations affect both the g- and ϵ -cleavage sites in APP, and because AICD has been suspected to play a role in the pathogenesis of Alzheimer's disease, we analyzed A β and AICD products generated from purified APP-C99 substrates carrying the FAD mutations listed in Figure 2a. All recombinant substrates were processed by purified γ -secretase (Supplementary Fig. 3). IP/MS analyses of the A β and AICD profiles revealed that amino acid substitutions clustered near the a-secretase (A692G, E693G, D694N) apparently do not alter the A β or AICD profiles, when compared to the control WT APP-C99 (Fig. 2b). In striking contrast, T714I and V175A mutations close to the transmembrane γ -cleavage site lead to drastically increased A β 42/40 and A β 38/40 ratios and to a total inhibition of the production of the shortest AICD50-99 (Fig. 2c). Interestingly, mutations closer to the ϵ -cleavage sites (I716V, V717F and L723P) are characterized by an intermediate phenotype as they do not alter A β profiles, but clearly lower the production of AICD50-99 (below detection levels for V717F and L723P) (Figs. 2b and c). Taken together, our data demonstrate that FAD-causing mutations in the transmembrane region of APP, but not close to the a-secretase site, affect both g- and/or ϵ -sites.

GSMs affect the g- but not the ϵ -cleavage of APP with FAD mutations.

To better understand the molecular mechanism by which GSMs shift A β production, as well as to verify whether GSMs can still lower A β 42 and have potential therapeutic benefits in patients with FAD, we decided to test GSM-1 on APP-C99 with FAD mutations. GSM-1 was chosen because it was more potent in our *in vitro* γ -secretase assay, and also because this compound did not affect e-cleavage in all cell-based experiments. First, 10 mM GSM-1 did not affect the levels of total cleavage products (compare AICD and A β levels in Supplementary Figs. 3a and 3b), providing further evidence that GSMs do not modify the overall yield of γ -secretase processing. Next, we found that GSM-1 conserved its A β 42-lowering and A β 38-raising properties on all tested mutant substrates (Fig. 3a). However, GSM-1 did not influence the e-cleavage and none of the mutants tested recovered the wild-type phenotype of nearly equimolar production of AICD49-99 and AICD50-99 (compare AICD profiles in Figs. 3b and 2c). Our findings provide new evidence that GSMs lower A β 42 production by interfering with the g-site in APP substrates with FAD mutations, strongly suggesting that none of the substituted amino-acid residues are crucial for the GSM-dependent modulation of APP processing by γ -secretase. They further demonstrate that GSMs fail to reverse the effects on AICD production of FAD-causing mutations in the transmembrane domain of APP.

Structural changes in the TM domain of APP with FAD-causing mutations

To support our results mechanistically, we used molecular dynamics (MD) simulations to further insights into the structural and dynamic properties of the transmembrane domain (TMD) of wild-type APP and three FAD-mutants (T714I, V715A and V717F) in the membrane environment. These mutations were chosen because they either affect both γ - and ϵ -cleavage sites (T714I, V715A), or exclusively the ϵ -site (V717F). The TM region of the APP-C99 NMR structure (PDB code 2LP1²⁷) was used and inserted into a phosphatidylcholine (POPC) lipid bilayer (Fig. 4a). MD simulations were carried out under *vacuo* physiological conditions (i.e., pH ~7, 1 atm, 300 K), mimicking as close as possible the conditions of the γ -secretase enzymatic assay used in our study. For all systems, the secondary structure of APP fluctuated only marginally. Consistent with the NMR observations, the N-terminus amphiphilic helix (Q686–N698) lies on the membrane surface, and a lysine belt (K687, K699, K724–726) anchors the TMD and N-terminus helices, respectively, in and onto the membrane (Fig. 4a). During simulation, the helices tilted with respect to the membrane surface. We found that the V717F mutation affected the most the orientation of the TMD, tilting it by 45 ± 5 degrees, while the wild type and T714I and V715A mutants remained more vertical (36 ± 5 , 38 ± 3 and 36 ± 3 degrees, respectively) (Fig. 4a and Supplementary Fig. 4a). These mutations further shifted the TMD perpendicularly to the membrane surface. When measuring the position of the A β 42 cleavage site (I712 - A713 peptide bond), T714I and V715A mutants shifted closer (2.9 ± 1.2 Å and 2.3 ± 0.9 Å, respectively) to the cytoplasmic interface compared to the WT, whereas V717F moved closer (-0.2 ± 1.3 Å) to the extracellular membrane layer (Fig. 4b). The combined effect of elevation and tilting perturbations of the APP TMD due to FAD-mutations is probably affecting the presentation to the enzyme of the γ -cleavage site at position 42, correlating quantitatively with the changes in A β 42/A β 40 ratio (Fig. 2). Moreover, the flexibility of the transmembrane domain, estimated based on the local bending of the helix²⁸, resulted to be much higher for T714I and V715A mutants, specifically at the G₂₉XXXGXXXG₃₇ region (Fig. 4c). Molecular docking and simulations of GSM-1 compound suggested the same region as the conserved drug-binding pocket (L701, G704 and G708 being the most direct interactions with GSM-1, Supplementary Fig. 4) for both the wild-type and mutant APP. Thus, GSM-1 effect on T714I and V715A FAD mutants might cause an attenuation of their flexibility close to the g-site. For all three mutants, the hydration of the ϵ -site at residue L720 was significantly decreased compared to the WT (Fig. 4d). In particular, for APP-WT one water molecule is in average within 3 Å of L720 backbone, while the water coordination number drops to less than 0.5 Å for V715A, and to nearly zero for T714I and V717F. This effect can be ascribed to the different water permeation properties of the APP C-terminus lysine belt combined with the perturbed helical orientation produced by the mutants. Together, our data support a model in which structural changes and, in particular, the

position of the γ -cleavage site relative to the membrane surface can potentially modulate A β production, while the hydration of L720 can affect the ϵ -cleavage site and the production of the two AICDs potentially implicated in signalling pathways and in the pathogenesis of AD.

DISCUSSION

Mutations in APP cause aggressive early-onset Alzheimer's disease by triggering abnormal production, accumulation and aggregation of A β peptides. The most severe pathological phenotype is observed for mutations occurring in the transmembrane region and in close vicinity of the A β 42 γ -cleavage site, leading to elevated A β 42/40 and A β 38/40 ratios (Fig. 2). Small chemical γ -secretase modulators (GSMs) are known to lower A β 42 and raise A β 38, without affecting the overall production of the A β peptides. In this study, we used second generation GSMs with a strong therapeutic potential (GSM-1 and E2012), in combination with APP substrates with FAD-causing mutations, (i) to better understand the molecular mechanisms of the γ -secretase-dependent, intramembrane and sequential cleavage of the substrate, (ii) to identify changes in APP-cleavage profiles, potentially implicated in the pathogenesis of AD, and (iii) to further support the therapeutic potential of GSMs to safely treat sporadic as well as early-onset familial AD. We further demonstrate here that GSMs represent an attractive scaffold for structure-activity relationship design (SAR) of potent disease-modifying drugs against AD. As they do not affect the processing of the Notch substrate (Fig. 1), GSMs are expected to circumvent clinical side effects observed with GSIs and attributed to impaired Notch processing¹⁷. Moreover GSMs do not alter AICD production (Fig. 1) and thus are not likely to influence the downstream cell signaling functions of APP, again underlying the huge therapeutic potential of these compounds in sporadic AD cases. Additionally, our results confirm altered g-site processing in APP-FAD mutations, associated with an increase of A β 42 and A β 38 and demonstrate that mutations located in close vicinity to the g-site do impede the production of the shorter ϵ -site counterpart AICD50-99. Mutations T714I and V715A display the strongest phenotype at both g- and ϵ -sites, with highest A β 42/40 and A β 38/40 ratios. Moreover, these mutations lead to the production of only one detectable AICD, suggesting a putative role of AICD in AD pathogenesis. These results might be explained with the sequential cleavage model of APP-CTF, in which the generation of A β 40 is dependent on the AICD50-99, while A β 42/38 is linked to AICD49-99¹⁴. Reduced AICD50-99 (as observed for some FAD mutants) is expected to be associated with less A β 40 and more A β 42/38, thus increasing the pathogenic ratio of A β 42/40. The combination of higher A β 42/40 ratio and the loss of AICD50-99 can explain the extreme aggressiveness of these APP mutations with regard to the onset of the disease. GSMs were still potent at the g-site of mutated APP substrates, but failed to rescue the phenotype at the ϵ -cleavage site (Fig. 3). These findings clearly demonstrate that GSMs can be used for the treatment of patients with both sporadic and genetic forms of AD. To better understand the effect of GSM binding, we docked the GSM-1 compound on the ensemble of wild-type and FAD-mutants (T714I, V715A and V717F) conformations generated by MD simulations. GSM-1 strongly binds to the previously identified G₂₉XXXGXXXG₃₇ motif (Supplementary Fig.4), further supporting our observation that amino acids close to the g-sites of APP are not required for GSMs activity, again underlying the therapeutic potential of these compounds²⁰.

Rationalizing these results with structural models based on molecular simulations of wild type and mutant APP substrates, we propose that the observed effects on A β and AICD production can be triggered by combined changes in the flexibility, tilting angle, and positioning of both g- and ϵ -sites in the membrane, along with the hydration levels at the ϵ -sites. Such structural changes, by potentially modifying the presentation of the substrates to the enzymatic complex, might affect the γ -secretase cleavage specificity. Consistent with this model, the elevated A β 38/40 ratio observed in mutants T714I and V715A can be explained by the vertical shift of the helix in the membrane bilayer, with the consequent presentation of position 38 instead of 40 to the γ -secretase cleavage site. The observed vertical shift in mutants (~2.5 Å) is in fact associated with a tilting of half of a helix turn (i.e., displacement of 2 amino acids). Interestingly,

A β 38 and A β 42 have been proposed to originate from the same sequential processing pathway, starting at the e-position 48, and directly linked to AICD49-99 production¹⁴. Consistent with this sequential model, we propose that the conformational changes in the T714I and V715A substrates, associated with reduced AICD50-99 levels, explain the preferential A β 38 and A β 42 production observed with these mutants. Interestingly, and apparently in contrast to the sequential cleavage model, the V717F and L723P mutants, although characterized by reduced AICD50-99 levels, do not show modifications in their A β profiles. This observation may reflect the existence of two or more different pathological molecular explanations, classifying FAD forms in different categories.

The nearly complete loss of AICD50-99 production in mutants T714I, V715A, V717F and L723P, can be attributed, according to our model, to higher solvation states around positions 49 and 50, with the latter becoming more exposed to water molecules. One can speculate that the increased water accessibility to the leucine 49 may favor the cleavage at this position, or inhibit the cleavage at valine 50 by preventing the interaction at position 50 of APP-C99 with the hydrophobic transmembrane domain 1 of the γ -secretase catalytic component PS1.

Together, our data reported here provide new insights into the therapeutic mechanism and potential of chemical γ -secretase modulators. They are further expected to improve our current understanding of the pathological mechanism of familial AD.

METHODS

γ -secretase inhibitors and modulators. γ -secretase inhibitors Compound E and GSI XIX were purchased from Merck Millipore (Darmstadt, Germany). The heteroaryl-type GSM E2012 (US2006/0004013) was obtained from Haoyuan Chemexpress Co. and the NSAID-type GSM-1 was synthesized at Syngene according to methods described in WO2006/043064.

Molecular Biology.

The pET21b vector encoding the cDNAs of human APP-C99-His was used as a template for the generation by PCR-amplification of all FAD APP-C99 mutants described in this study. The constructs were expressed in BL21(DE3) cells induced with 1 mM isopropyl- β -D-thiogalactopyranoside (IPTG) for 4 h at optic density 600 nm = 0.8. For the purification of APP-C99 mutants, harvested bacteria were lysed with 10 mM Tris pH 7.0, 150 mM NaCl, 1 % Triton X-100, protease inhibitor cocktail (PI; Roche, Rotkreuz, Switzerland), and passed through a high-pressure homogenizer (Emulsiflex-C5; Avestin, Inc., Mannheim, Germany). The obtained lysates were spun down at 5000 x g for 30 min and supernatants were incubated overnight with Ni-NTA agarose beads (Invitrogen, Basel, Switzerland) at 4 °C. Bound proteins were eluted in 1 % NP-40 containing 250 mM imidazole (pH 7.8), analyzed by Coomassie stained sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and quantified by Bicinchoninic Acid (BCA; Pierce, Rockford, IL, USA). The normalization was validated by WB. Recombinant APP-C100-Flag and Notch N100-Flag substrates were also expressed in BL21(DE3) cells, but 100 mM glycine (pH 2.7)/1% NP40 was used for elution from M2-agarose beads.

Western blotting and antibodies.

Samples from γ -secretase activity assays were run on 4–12 % Bis-Tris gels and transferred onto PVDF membranes to detect Ab, Flag-tagged APP- or Notch-intracellular domain (AICD-Flag and NICD-Flag) with Ab-specific 6E10 (1:1000; Covance, Berkeley, CA, USA) and Flag-specific M2 antibodies. Anti-

mouse/rabbit/rat IgG conjugated to Alexa 680 were purchased from Invitrogen and the Odyssey infrared imaging system (LICOR) was used to detect the fluorescent signal.

Whole cell extracts prepared in 50 mM HEPES buffer containing 1 % NP40 were separated by electrophoresis on NuPAGE® Novex® 4-12 % Bis-Tris Gels (Invitrogen, Carlsbad, CA, USA) for SDS-PAGE analysis transferred onto nitrocellulose membranes, and probed with antibodies against APP-CTFs (Sigma-Aldrich GmbH, Buchs, Switzerland.). For detection of NICD, membranes were blocked with 5 % milk and incubated overnight with primary antibody Val1744 anti-Notch (Cell Signaling, Bioconcept, Allschwil, Switzerland) at 4 °C. Membranes were washed with 1x TBST (1x TBS + 0.1 % Tween 20) for 15 min and incubated with HRP-conjugated secondary antibodies for 1 h at room temperature. Signal was detected with Super Signal West chemiluminescent substrate (Thermo Scientific, catalog number 34077).

γ-secretase purification. γ-secretase used in this study was purified from 10 liters of PB suspension cultures by following the protocol described by Alattia et al. (submitted for publication).

γ-secretase activity assays. γ-secretase assays using the recombinant APP- or Notch-based substrates were performed as previously reported²⁹. Briefly, highly-purified γ-secretase were solubilized in 0.2 % (wt/vol) CHAPSO, 50 mM HEPES (pH 7.0), 150 mM NaCl, 5 mM MgCl₂ and 5 mM CaCl₂, and incubated at 37 °C for 4 h with 1 mM substrate, 0.1 % (wt/vol) PC, 0.025 % (wt/vol) PE. The resulting products, APP intracellular domain AICD-Flag, AICD-His, Ab, and Notch intracellular domain NICD-Flag were detected with above-described antibodies, or analyzed by mass spectrometry.

Immunoprecipitation-mass spectrometry (IP-MS) analysis of Aβ and AICD. Aβ and AICD generated in γ-secretase *in vitro* assays with purified substrates were analyzed as previously described²⁹. Briefly, Aβ was IP'd overnight using anti-Aβ antibody 4G8 (Covance, Berkeley, CA, USA) and protein G coupled to agarose resin (Roche Applied Science, Penzberg, Germany). For AICD or NICD detection, 12 mL 10 % Triton-X100 solution was added after the enzymatic reaction and incubated for 20 min at 55 °C prior to overnight immunoprecipitation at 4 °C with M2-anti Flag resin (Invitrogen, Carlsbad, CA, USA), or Protein G (Roche, Mannheim, Germany), coupled with anti-C-terminal AICD antibody. Aβ and AICD-Flag were eluted with 1:20:20 (v:v:v) 1 % (vol/vol) trifluoroacetic acid:acetonitrile:H₂O, equally mixed with saturated CHCA (α-cyano 4-hydroxy cinnaminic acid), and analyzed by MALDI-TOF mass spectrometry in reflectron mode on an ABI 4800 MALDI-TOF/TOF mass spectrometer (Applied Biosystems, Carlsbad, CA, USA). Molecular masses were accurately measured and searched against amino acid sequences of human APP-C99 with addition of a methionine residue at the N-terminus and a Flag tag sequence (C100Flag) or a His tag sequence (C100His) at the C-terminus, or a A692G, E693G, D654N, T714I, V715A, I716V, V717F and L723P substitution.

Molecular dynamics simulations and molecular docking. We used molecular dynamics (MD) simulations to characterize the TM domain of WT and APP with FAD mutations. In a first step, the NMR structure (PDB code 2LP1²⁷) was inserted and equilibrated in a 60 x 60 Å² Palmitoyl Oleoly Phosphatidyl Choline (POPC) patch³⁰ to relax the structure in a phospholipid bilayer. The system was solvated in 60 x 60 x 100 Å² water box, neutralized through the addition of NaCl at a concentration of 150 mM. After 30 ns, the equilibrated structure was used as initial seed to build the other three mutants (V714A, T715I and V717F) and WT-ligand complex. GSM-1 ligand was parameterized using CGenFF and ParamChem 0.9.6³¹. All MD simulations were performed using NAMD 2.8 engine, with the CHARMM27 force field, including CMAP corrections for the protein and CHARMM36 for the POPC membrane^{32,33}. TIP3P water parameterization was used to describe the water molecules. The periodic electrostatic interactions were computed using the particle-mesh Ewald (PME) summation with a grid spacing smaller than 1 Å. All systems were equilibrated for 30 ns at 300K. Free molecular dynamics were performed up to 100 ns with a 2 fs integration time step using the RATTLE algorithm applied to all bonds. Constant temperature (300 K) was imposed by using Langevin dynamics, with damping coefficient of 1.0 ps. Constant pressure of 1 atm was maintained with a Langevin piston dynamics, 200 fs decay period and 50 fs time constant.

GSM-1 was docked on WT and APP mutants using SWISSDock^{34,35} and BSP-SLIM³⁶; independently both approaches produced similar poses. Since these algorithms do not account for the presence of the membrane, we discarded all unphysical docking poses, e.g. the GSM-1 hydroxyl being buried in the hydrophobic membrane core. The conformation with the lowest energy was then selected for each system.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation (to P.C.F., grant 31003A_134938/1, to M.D.P. grant 200020_138013), the Swiss National Centres of Competence in Research, Neural Plasticity and Repair (to P.C.F.), the Strauss foundation (to P.C.F.), and an EPFL-Merck-Serono grant (to J.R.A. and M.D.). The authors thank M. Moniatte (Proteomics Core Facility, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland) for technical assistance with mass spectrometry, Laurence Anderes for the Notch assay in SUP-T1 cells, and M.S. Wolfe for the T20 cell line. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

M.D., L.R., J.R.A., J.H and A.F. performed experiments; T.L. performed and analyzed with M.D.P. molecular simulations; I.H., F.R., D.B., M.D.P. and P.C.F. designed and supervised the study; M.D. and P.C.F. wrote the manuscript. All authors edited the paper.

COMPETING FINANCIAL INTERESTS

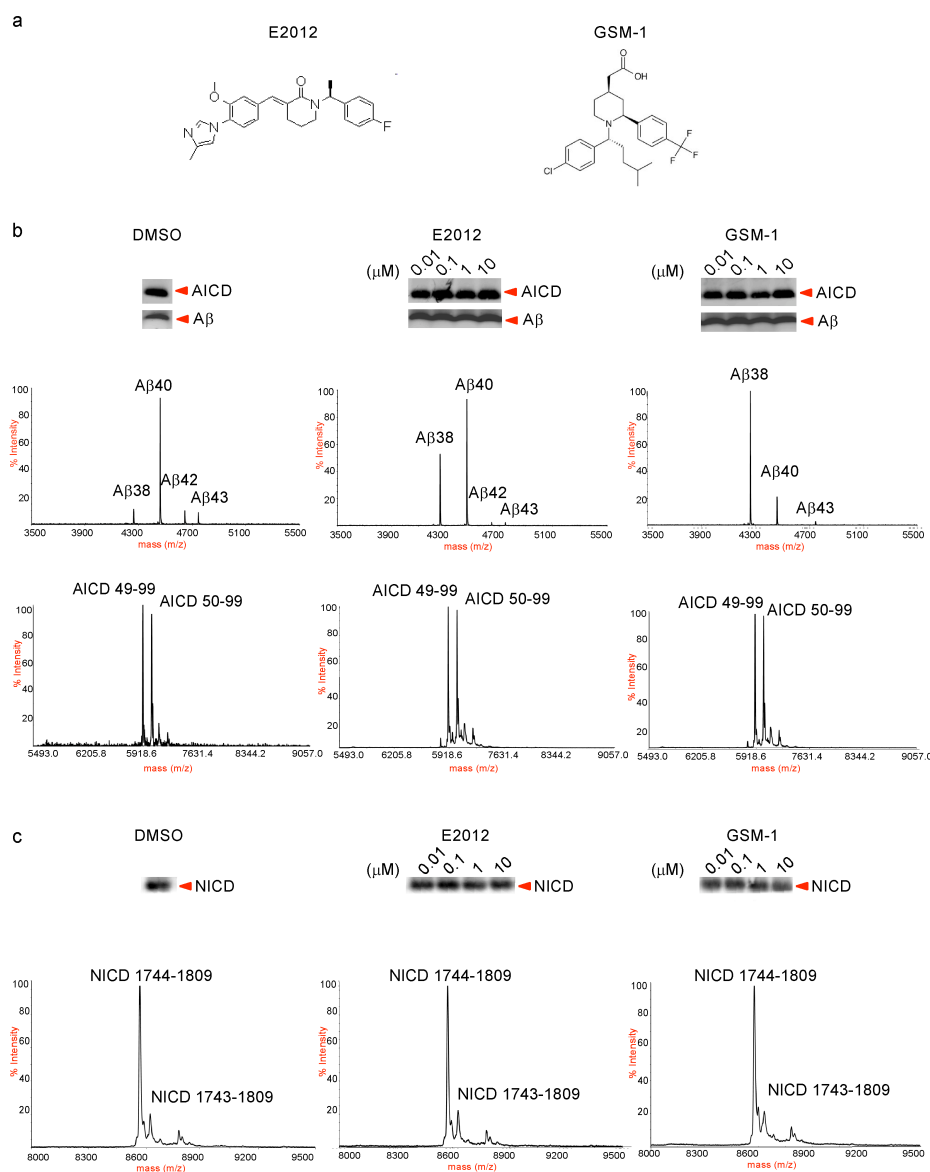
D.B. and I.H. are full time employees of Merck Serono S.A. Other authors declare having no competing financial interests.

REFERENCES

1. Masters, C.L. et al. Amyloid plaque core protein in Alzheimer disease and Down syndrome. *Proc Natl Acad Sci U S A* **82**, 4245-9 (1985).
2. Robakis, N.K. et al. Chromosome 21q21 sublocalisation of gene encoding beta-amyloid peptide in cerebral vessels and neuritic (senile) plaques of people with Alzheimer disease and Down syndrome. *Lancet* **1**, 384-5 (1987).
3. Goate, A.M. et al. Predisposing locus for Alzheimer's disease on chromosome 21. *Lancet* **1**, 352-5 (1989).
4. Sherrington, R. et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* **375**, 754-60 (1995).
5. Haass, C. et al. Normal cellular processing of the beta-amyloid precursor protein results in the secretion of the amyloid beta peptide and related molecules. *Ann N Y Acad Sci* **695**, 109-16 (1993).
6. De Strooper, B. et al. A presenilin-1-dependent gamma-secretase-like protease mediates release of Notch intracellular domain. *Nature* **398**, 518-22 (1999).
7. Bot, N., Schweizer, C., Ben Halima, S. & Fraering, P.C. Processing of the synaptic cell adhesion molecule neurexin-3beta by Alzheimer disease alpha- and gamma-secretases. *J Biol Chem* **286**, 2762-73 (2011).
8. Uemura, K. et al. Characterization of sequential N-cadherin cleavage by ADAM10 and PS1. *Neuroscience Letters* **402**, 278-283 (2006).
9. Ohki, Y. et al. Phenylpiperidine-type gamma-secretase modulators target the transmembrane domain 1 of presenilin 1. *EMBO J* **30**, 4815-24 (2011).
10. Lazarov, V.K. et al. Electron microscopic structure of purified, active gamma-secretase reveals an aqueous intramembrane chamber and two pores. *Proc Natl Acad Sci U S A* **103**, 6889-94 (2006).
11. Cacquevel, M. et al. Rapid purification of active gamma-secretase, an intramembrane protease implicated in Alzheimer's disease. *J Neurochem* **104**, 210-20 (2008).
12. Chavez-Gutierrez, L. et al. The mechanism of gamma-Secretase dysfunction in familial Alzheimer disease. *EMBO J* **31**, 2261-74 (2012).
13. Haass, C. et al. The Swedish mutation causes early-onset Alzheimer's disease by [beta]-secretase cleavage within the secretory pathway. *Nat Med* **1**, 1291-1296 (1995).
14. Sato, T. et al. Potential link between amyloid beta-protein 42 and C-terminal fragment gamma 49-99 of beta-amyloid precursor protein. *J Biol Chem* **278**, 24294-301 (2003).
15. Eckman, C.B. et al. A new pathogenic mutation in the APP gene (I716V) increases the relative proportion of A beta 42(43). *Hum Mol Genet* **6**, 2087-9 (1997).
16. Kumar-Singh, S. et al. Nonfibrillar diffuse amyloid deposition due to a gamma(42)-secretase site mutation points to an essential role for N-truncated A beta(42) in Alzheimer's disease. *Hum Mol Genet* **9**, 2589-98 (2000).
17. Schor, N.F. What the halted phase III γ -secretase inhibitor trial may (or may not) be telling us. *Annals of Neurology* **69**, 237-239 (2011).
18. Weggen, S. et al. A subset of NSAIDs lower amyloidogenic Abeta42 independently of cyclooxygenase activity. *Nature* **414**, 212-6 (2001).
19. Eriksen, J.L. et al. NSAIDs and enantiomers of flurbiprofen target gamma-secretase and lower Abeta 42 in vivo. *J Clin Invest* **112**, 440-9 (2003).
20. Kukar, T.L. et al. Substrate-targeting gamma-secretase modulators. *Nature* **453**, 925-9 (2008).

21. Crump, C.J. et al. Piperidine acetic acid based gamma-secretase modulators directly bind to Presenilin-1. *ACS Chem Neurosci* **2**, 705-710 (2011).
22. Mangialasche, F., Solomon, A., Winblad, B., Mecocci, P. & Kivipelto, M. Alzheimer's disease: clinical trials and drug development. *Lancet Neurol* **9**, 702-16 (2010).
23. Bulic, B. et al. Chemical Biology, Molecular Mechanism and Clinical Perspective of gamma-Secretase Modulators in Alzheimer's Disease. *Curr Neuropharmacol* **9**, 598-622 (2011).
24. O'Neil, J. et al. FBW7 mutations in leukemic cells mediate NOTCH pathway activation and resistance to gamma-secretase inhibitors. *J Exp Med* **204**, 1813-24 (2007).
25. Liao, Y.-F., Wang, B.-J., Cheng, H.-T., Kuo, L.-H. & Wolfe, M.S. Tumor Necrosis Factor- α , Interleukin-1 β , and Interferon- γ Stimulate gamma-Secretase-mediated Cleavage of Amyloid Precursor Protein through a JNK-dependent MAPK Pathway. *Journal of Biological Chemistry* **279**, 49523-49532 (2004).
26. Walsh, D.M., Hartley, D.M., Condron, M.M., Selkoe, D.J. & Teplow, D.B. In vitro studies of amyloid beta-protein fibril assembly and toxicity provide clues to the aetiology of Flemish variant (Ala692-->Gly) Alzheimer's disease. *Biochem J* **355**, 869-77 (2001).
27. Barrett, P.J. et al. The amyloid precursor protein has a flexible transmembrane domain and binds cholesterol. *Science* **336**, 1168-71 (2012).
28. Dahl, A.C., Chavent, M. & Sansom, M.S. Bendix: intuitive helix geometry analysis and abstraction. *Bioinformatics* **28**, 2193-4 (2012).
29. Alattia, J.R. et al. Mercury is a direct and potent gamma-secretase inhibitor affecting Notch processing and development in Drosophila. *FASEB J* **25**, 2287-95 (2011).
30. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-8, 27-8 (1996).
31. Vanommeslaeghe, K. et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* **31**, 671-90 (2010).
32. Phillips, J.C. et al. Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-802 (2005).
33. Klauda, J.B. et al. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J Phys Chem B* **114**, 7830-43 (2010).
34. Grosdidier, A., Zoete, V. & Michielin, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res* **39**, W270-7 (2011).
35. Grosdidier, A., Zoete, V. & Michielin, O. Fast docking using the CHARMM force field with EADock DSS. *J Comput Chem* (2011).
36. Lee, H.S. & Zhang, Y. BSP-SLIM: a blind low-resolution ligand-protein docking approach using predicted protein structures. *Proteins* **80**, 93-110 (2012).
37. Jeffries, S., Robbins, D.J. & Capobianco, A.J. Characterization of a high-molecular-weight Notch complex in the nucleus of Notch(ic)-transformed RKE cells and in a human T-cell leukemia cell line. *Mol Cell Biol* **22**, 3927-41 (2002).

FIGURE LEGENDS



Figures 1. GSMs alter g-site processing, but do not modulate ϵ -site on APP and Notch *in vitro*. **a)** Chemical structure of clinically tested small molecules NSAID-based phenylpiperidine-type GSM-1 and bridged aromatics E2012. **b)** GSMs do not inhibit the overall cell-free processing of APP as shown by WB analysis of A β and AICD at concentrations ranging from low nanomolar to low micromolar. Following immunoprecipitation, analysis by MALDI-TOF demonstrated the efficacy of GSMs on the g-site, where they shifted the production towards shorter A β 38 and lowered A β 42. Even at concentrations as high as 10 mM, GSMs do not alter the ϵ -site as evidenced by the IP/MS on the AICD products, AICD49-99 and AICD50-99. **c)** GSMs show no inhibition on Notch intracellular signaling products. *In vitro* NICD formation detected by WB is unaltered. The ratio of the two NICD products, NICD1744-1809 and NICD1743-1809, detected after IP/MS, also remained without change in the presence of 10 mM of both compounds.

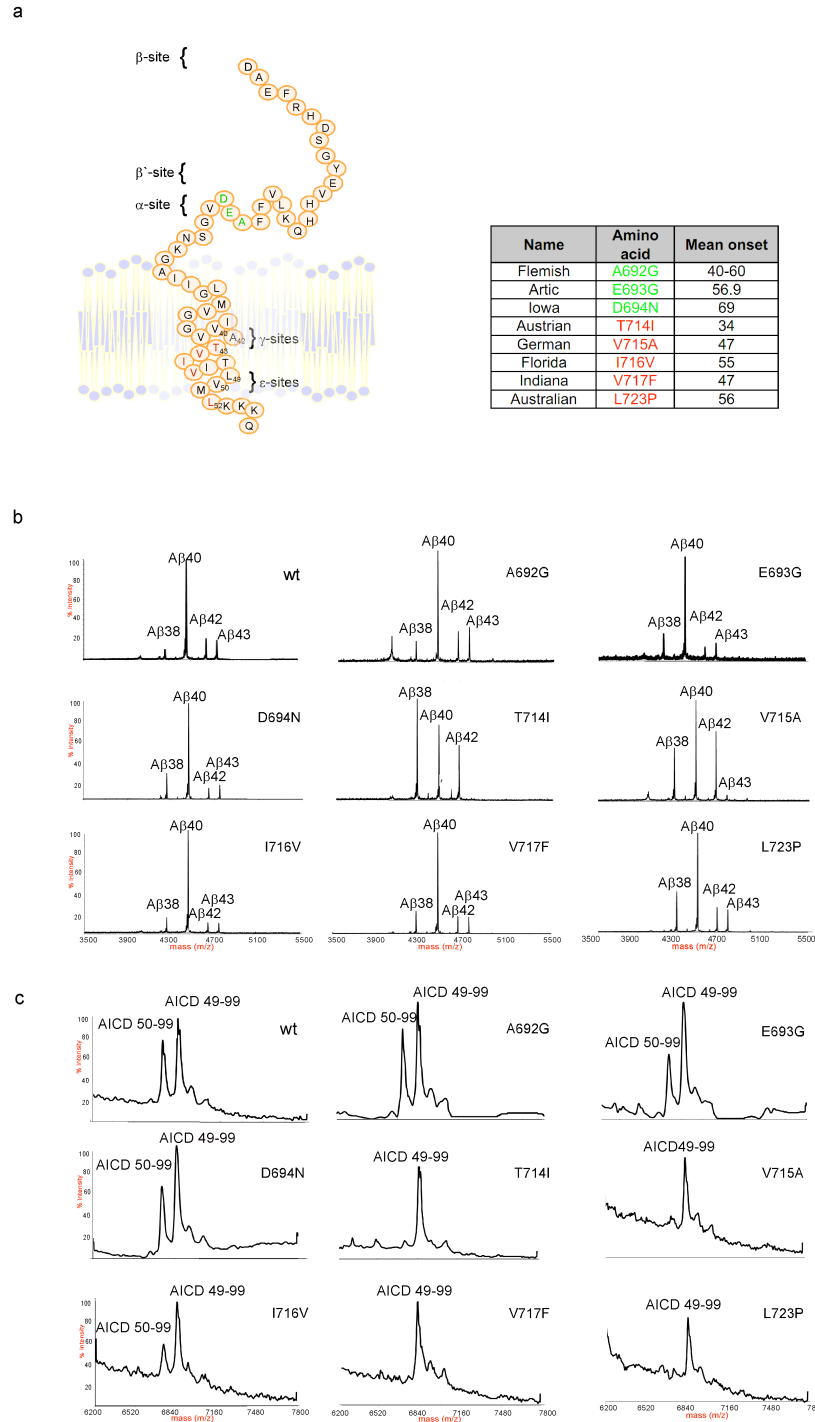


Figure 2. Pathological mutations in APP cause altered processing at both α - and γ -sites. **a)** FAD-APP mutations differ by the onset, but are all clustered close to the α -secretase and γ -secretase cleavage sites, as represented schematically in the left panel (color coding: α -secretase site mutations in green; γ -secretase site mutations in red). Different mean onsets are summarized in the right table. The earliest onset is reported for Austrian and German families at the age of early 50s. **b)** IP/MS profiles of A β

generated from purified FAD-APP-C99 substrates show that mutations A692G, E693G, D694N, I716V, V717F and L723P have similar and WT-like A β 38/40 and A β 42/40 ratios. In contrast, T714I and V715A mutations in the transmembrane region led to a drastic increase in both A β 38 and A β 42 and A β 38/40 and A β 42/40 ratios. **c)** AICD profiles from the same reactions as analyzed by IP/MS. FAD-APP-C99 mutations close to the α -secretase site produce AICD49-99 and AICD50-99 in nearly equimolar ratios. Amino acid substitutions in close vicinity of the g-sites in the TMD strongly affect the yield of shorter AICD50-99.

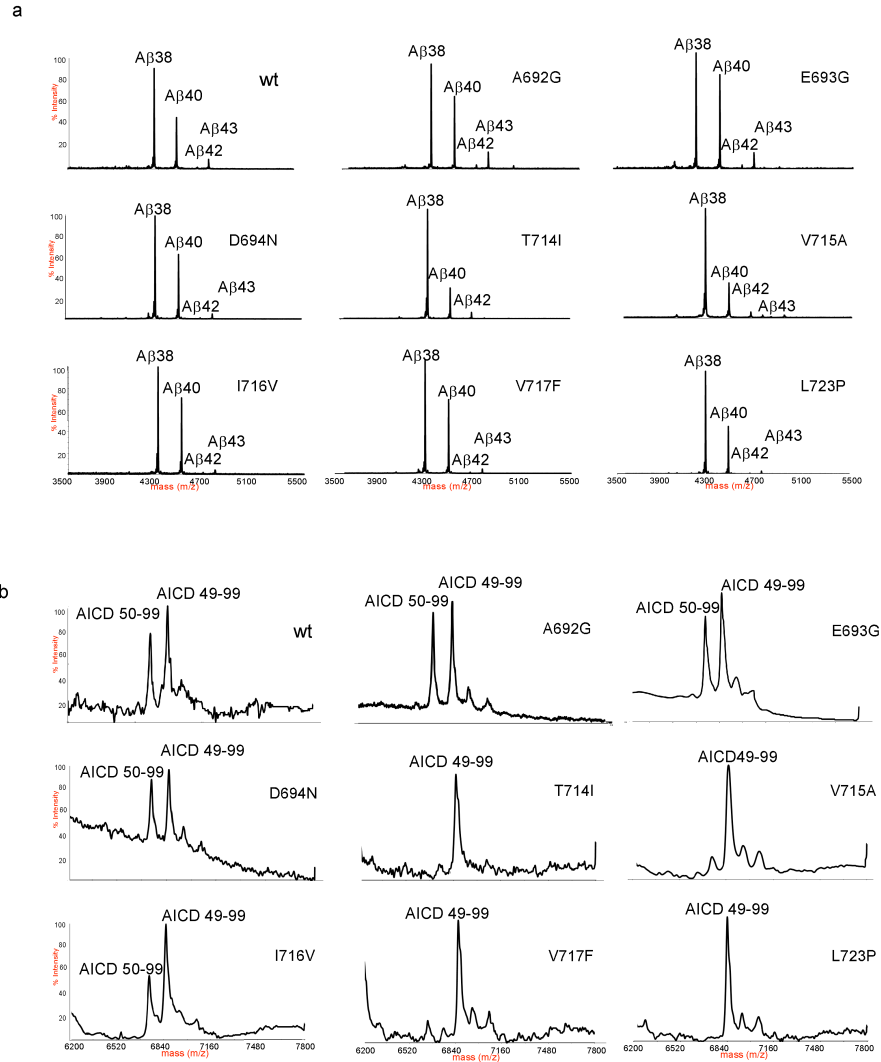


Figure 3. GSMs potentially shift the production at g-, but not at ϵ -sites in APP with FAD mutations. **a)** IP/MS profiles of A β species produced from purified APP substrates with FAD mutations detected by IP/MS demonstrate the efficacy of GSMs towards the g-site processing. For all FAD-APP mutants tested, A β 38/40 and A β 42/40 ratios are strongly affected by 10 mM of GSM-1. Mutations T714I and V715A had the most elevated A β 38/40, implying that amino acids between APP g- and ϵ -sites are not essential for the GSM mechanism of action. **b)** IP/MS profiles of the ϵ -site products from the same reactions provide supplementary evidence that GSMs do not influence the ϵ -sites in APP. Even high concentrations of GSM-1 fail to restore the loss of shorter AICD50-99 in mutants T714I, V715A, V717F and L723P.

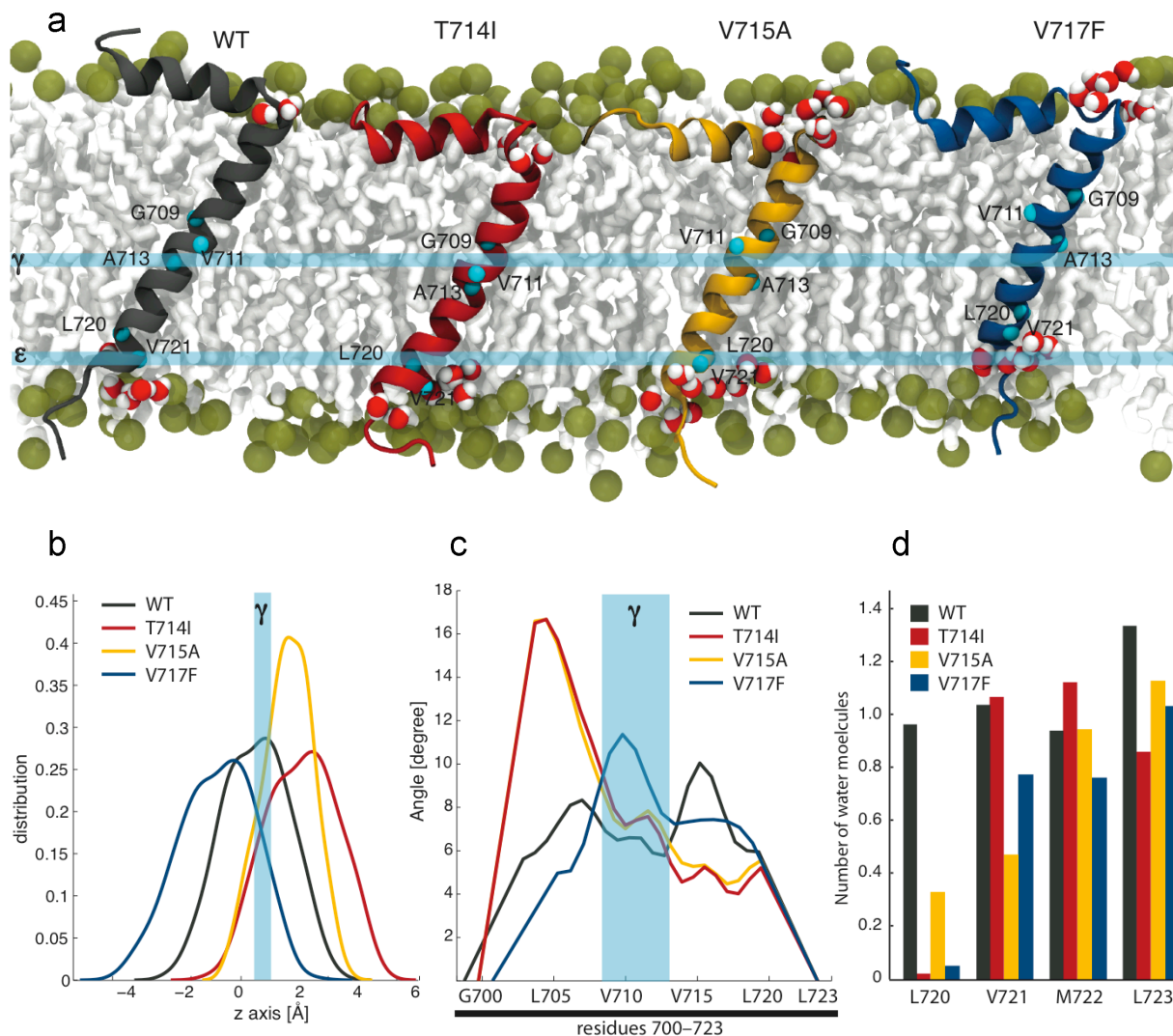


Figure 4. Structural and dynamic models of the TM domain in WT and FAD-causing APP mutants. **a)** The NMR structure of WT APP (PDB code: 2LP1²⁷) and three FAD-causing mutants, namely T714I, V715A and V717F, were inserted in a POPC membrane bilayer and simulated with molecular dynamics. The g-sites G709, V711 and A713 and ε-sites L720 and V721 positions are indicated by blue spheres. **b)** The spatial distribution of Aβ42 cleavage site is reported for the four systems. The distribution is calculated along the axis orthogonal to the membrane bilayer surface and the reference putative position of the g-site in wild-type conditions is highlighted with the blue area. **c)** The local bending of the APP helix is computed for the TM domain (with a standard error of 2.8 degrees). **d)** The coordination number of water molecules around the relevant residues at the ε-site is reported for the WT and mutants.

In depth analysis of different Hyaline Fibromatosis Syndrome frameshift mutations at the same site reveal the necessity of personalized therapy

Shixu E. Yan¹, Thomas Lemmin², Suzanne Salvi¹, Ekkehart Lausch³, Andrea Superti-Furga⁴, Dariusz Rokicki⁵, Matteo Dal Peraro² and F. Gisou van der Goot^{1*}

¹ *Global Health Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), Station 19, CH-1015 Lausanne, Switzerland*

² *Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Station 15, CH-1015 Lausanne, Switzerland*

³ *Department of Pediatrics, University of Freiburg, Freiburg, Germany* ⁴ *Division of Molecular Pediatrics, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Switzerland*

⁵ *Division of Inborn Errors of Metabolism, Children's Memorial Health Institute, Warsaw, Poland*

* Corresponding author

Ecole Polytechnique Fédérale de Lausanne Institute of Global Health
Station 15

1015 Lausanne Switzerland

Tel: (41) 021 693 1791

Fax: (41) 021 693 9538

Contact: gisou.vandergoot@epfl.ch

ABSTRACT

Hyaline Fibromatosis Syndrome (HFS) is an autosomal recessive disease caused by mutations in *cmg2*, a gene involved in extracellular matrix homeostasis. Sixty percent of patients carry a frameshift mutation at a mutational hotspot in exon 13, generally due to the insertion of one or two bases. We show in patient cells that these mutations lead to low *cmg2* mRNA and undetectable protein levels. Ectopic expression of the proteins encoded by the mutated genes reveals that a 2 base insertion leads to the synthesis of a protein that is rapidly targeted to degradation due to the modified structure of the cytosolic tail, which instead of being hydrophilic and highly disordered as in wild type CMG2, is folded and exposes hydrophobic patches. In contrast, a one base insertion leads a truncated protein that properly localizes to the plasma membrane and retains partial function. We next show that targeting the non-sense mediated mRNA decay pathway leads to a rescue of CMG2 protein in cells from patients carrying 1 base insertion but not in those carrying 2 base insertions.

INTRODUCTION

Hyaline Fibromatosis Syndrome (HFS) is a rare autosomal recessive disease caused by mutations in Capillary Morphogenesis Gene 2 (*cmg2*) (Dowling et al, 2003; Hanks et al, 2003). The disease is progressive, disfiguring and disabling. As suggested by the name, it is characterized by the accumulation of amorphous, unidentified hyaline material in the skin and other organs of patients (Shieh et al, 2006; Urbina et al, 2004), which could be either a direct effect or a secondary consequence due to the alterations in the basal membrane of capillary vessels (Stucki et al, 2001). These non-cancerous tissue proliferations are the most outstanding external hallmark of the patients (Deuquet et al, 2011b). In addition, patients usually present papular skin lesions, gingival hyperplasia, and joint contractures (Urbina et al, 2004). Two allelic forms are now subsumed in HFS: Infantile Systemic Hyalinosis (ISH), the more severe form, whose patients have very early onset even at birth and suffer from severe diarrhea, recurrent infection and malnutrition that lead to death; and the milder form, Juvenile Hyaline Fibromatosis (JHF), for which afflicted individuals reach adulthood even though highly incapacitated by the cutaneous tumors (Deuquet et al, 2009). Better pediatric care in severely affected infants allows some of them to survive infancy; these infants grow to develop the full-blown picture of JHF. Molecular results have confirmed that these disorders form a continuous phenotypic spectrum determined at least partially by the combination of

Early-onset Alzheimer's disease mutations in APP, but not γ -secretase modulators, affect epsilon-cleavage-dependent AICD production.

Mitko Dimitrov¹, Jean-René Alattia¹, Thomas Lemmin², Lehal Rajwinder³, Andrzej Fligier¹, Jemila Houacine¹, Ishrut Hussain⁴, Freddy Radtke³, Matteo Dal Peraro², Dirk Beher⁴, and Patrick C. Fraering^{1*}

¹ Brain Mind Institute and School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH1015 Lausanne, Switzerland

²Institute of Bioengineering and School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

³Swiss Institute for Experimental Cancer Research and School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH1015 Lausanne, Switzerland

⁴Global Research and Early Development, Merck Serono SA, 9 Chemin des Mines, 1202 Geneva, Switzerland

* Correspondence should be addressed to Patrick C. Fraering, Brain Mind Institute and School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH1015 Lausanne, Switzerland. E-mail: patrick.fraering@epfl.ch

ABSTRACT

Pathological amino acid substitutions in the amyloid precursor protein (APP) and chemical γ -secretase modulators (GSMs) affect the processing of APP by the γ -secretase complex, and the production of the amyloid-beta peptide A β 42, the accumulation of which is considered causative of Alzheimer's disease. We demonstrate here that mutations in the transmembrane domain of APP causing aggressive early-onset familial Alzheimer's disease (FAD) affect both γ - and ϵ -cleavage sites, by raising the A β 42/40 ratio and inhibiting the production of AICD50-99, one of the two physiological APP intracellular domains (ICDs). This is in sharp contrast to GSMs, which shift A β 42 production towards the shorter A β 38, but unequivocally spare the ϵ -site and APP- and Notch-ICDs production. Molecular simulations suggest that FAD mutations modulate the flexibility of the APP transmembrane domain and the presentation of its γ -site, modifying at the same time the solvation of the ϵ -site.

INTRODUCTION

The main pathological hallmark of Alzheimer's disease (AD) is the formation of toxic amyloid fibers that progressively accumulate into plaques, which leads to degeneration of the neighboring neuronal cells¹. Further biochemical studies identified the main protein component of plaques to be a short fragment of APP, the 42 amino acid long peptide A β 42². Moreover, dominantly inherited mutations causing Familial Alzheimer's disease (FAD) are found in APP as well as in the PSEN1 and PSEN2^{3,4} genes encoding for the catalytic subunit of the γ -secretase complex, an intramembrane-cleaving protease which cleaves APP in the plasma membrane and generates the A β peptides⁵. Together, these findings clearly demonstrated the implication of APP, γ -secretase and A β in both the sporadic and genetic forms of AD. They further highlighted the γ -secretase-dependent APP processing as an attractive target for drug development.

In addition to APP, γ -secretase cleaves various other substrates regulating biological functions including development and stem cell differentiation. This unusual proteolytic complex indeed plays an important role in cell signaling pathways by processing different type-I transmembrane cell receptors including the Notch receptors, the Insulin receptor, and the Growth hormone receptor⁶. In the brain, it maintains synaptic contacts by regulating cleavage of neuronal adhesion molecules like Cadherins and Neurexins^{7,8}. With the continuously growing number of substrates the need for understanding the exact molecular mechanism of this particular proteolytic processing arises.

γ -Secretase is composed of four subunits, presenilin (PS), Aph-1, Pen-2 and Nicastrin (NCT). The complex contains in total 19 α -helices embedded in the membrane bilayer, which allows the overall structure to adopt different conformations upon binding with the substrate or transition-state analog inhibitors⁹. Moreover, the catalytic aspartate residues (Asp257/385) in PS1 are buried into the lipid bilayer and surrounded by hydrophobic helices and the cryo-electron microscopy density maps suggested the existence of two pores in the globular structure of γ -secretase, which can explain the entrance of water molecules necessary for the hydrolysis of the peptide bond¹⁰. Yet, a rigorous biochemical characterization of γ -secretase complexes with FAD mutations revealed that they all represent a loss of enzymatic function, but they ultimately lead to altered A β production and elevated A β 42/40 ratio^{11,12}. Interestingly, FAD mutations in APP can show substrate cleavage patterns that differ from those of mutations in γ -secretase. The large extracellular domain of APP is first removed by b-secretase in the vicinity of the plasma membrane. Gain-of-function mutations in this b-site cause elevated levels of total Ab¹³. γ -secretase processes the remaining 99-amino-acid APP-C terminal fragment (APP-C99) at ϵ -sites close to inner layer of the membrane, and releases 49- and 50-amino-acid long APP- intracellular domains (AICD50-99 and AICD49-99). After this step, γ -secretase sequentially continues to trim the residual transmembrane stub, finally producing A β 42, A β 40 and A β 38 peptides¹⁴. Mutations surrounding the ϵ -sites either increase the production of A β 42, which is more neurotoxic compared to the shorter and less

fibrillogenic A β 40, or lead to elevated total amounts of the A β product without affecting A β ratios^{12,15}. Amino acid substitutions in the g-site lead to high A β 42/40 ratios and are characterized by the most prominent and earliest phenotypes of FAD¹⁶.

Inhibiting the γ -secretase activity to reduce A β production became an attractive therapeutic strategy to treat AD. However, γ -secretase inhibitors (GSIs) tested in clinical trials not only did not meet their initial clinical endpoints, but also manifested severe off-target effects due to impaired Notch signaling, leading to gastrointestinal bleeding and increased risk of skin cancer¹⁷. Consequently, a concept of selective modulation of the γ -secretase activity has emerged. A subset of non-steroidal anti-inflammatory drugs (NSAIDs) were the first Notch-sparing small molecules shown to selectively lower A β 42 and subsequently raise A β 38, both *in vitro* and *in vivo*^{18,19}. These γ -secretase modulators (GSMs) were initially shown to bind to APP, although some more recent studies demonstrated that they also can target γ -secretase^{20,21}. Although the precise molecular mechanism of action remains poorly understood, a number of these molecules have advanced to clinical trials²².

In this report, we demonstrate that second generation GSMs, including the clinically tested E2012, effectively lower the production of toxic A β 42, without affecting the e-cleavage sites of both APP and Notch substrates. In contrast, we show that the T714I, V715A, V717F and L723P FAD-APP mutations in the transmembrane (TM) region block the production of AICD50-99, the 49-amino-acid long AICD, without affecting the total amount of AICD. Moreover, mutations T714I and V715A elevated the A β 42/40 and A β 38/40 ratios, in contrast to I716V, V717F and L723P for which these profiles were comparable to those obtained with wild type APP. GSMs affected all mutants at their γ -cleavage site, with the strongest effect on the A β 38/40 and A β 42/40 ratios in T714I and V715A, suggesting that these amino acid substitutions are important for the generation of A β 42 and A β 38. In addition, we found that GSMs do not reverse the loss of AICD50-99 production caused by FAD mutations. We rationalize these findings using molecular dynamics simulations of the APP TM segment embedded in the lipid bilayer, and propose that FAD-APP mutants and possibly GSMs affect the γ -cleavage site by changing the flexibility, tilt angle and position of the substrate in the membrane. GSMs fail to recover e-cleavage, most likely because they affect the substrate-enzyme interface at a position which is distant from the ϵ -site and closer to the outer layer of the membrane. Finally, FAD mutants had altered solvation at the e-cleavage site, which can explain the different product lines generated after the first step of the processing of APP.

RESULTS

GSMs affect the g- but not the e-cleavage of APP and Notch.

We first investigated the effect of GSM-1 and E2012, two representative molecules of the two main classes of second generation GSMs: NSAID-based phenylpiperidine-type compounds and non-NSAID bridged aromatics, respectively (Fig. 1a). Although these compounds have different structures, they both modulate γ -secretase activity at low nanomolar concentrations in cells²³. In enzymatic *in vitro* assays using highly purified γ -secretase and recombinant APP-C100-Flag substrate, we first show that concentrations several fold higher than the cellular IC₅₀ concentrations of GSM-1 and E2012 do not affect the overall amounts of A β and AICD, when compared to control levels (Fig. 1b, upper panels). We next investigated in detail the effect of these molecules on all reaction products. For both compounds tested at 10 mM, immunoprecipitation with 4G8 anti-A β antibody, combined with MALDI-TOF spectrometry (IP/MS) revealed a robust increase in A β 38, associated with a decrease in A β 42 (Fig.1b, middle panels). At this concentration GSM-1 showed a stronger effect than E2012 on A β 42/40 and A β 38/40 ratios, while both compounds lowered A β 43. IP/MS analyses of the C-terminal cleavage products revealed an unchanged equimolar production of the two AICDs (Fig. 1b, lower panels). Together, these results clearly demonstrate that GSMs affect the APP γ -cleavage sites, while completely sparing the ϵ -sites. Because of

the clinical relevance of GSMs and the Notch signaling pathway, we further questioned whether GSMs could interfere with the Notch intracellular domain (NICD) production. As shown in Figure 1c, GSMs did not influence total NICD production and did not affect the NICD cleavage profiles. Our *in vitro* findings were further confirmed in cell-based assays. First, intracellular accumulation of endogenous γ -secretase substrates APP-C83 and APP-C99, which is a common characteristic among known γ -secretase substrates in response to γ -secretase inhibition, was assessed in HEK293T cells stably overexpressing APP with the FAD Swedish mutation KM670/671NL. In contrast to the control GSI, 1 and 10 mM E2012 or GSM-1 did not affect the levels of APP-CTFs (Supplementary Fig. 1a). Next, treatments with E2012 and GSM-1 up to 1 mM in a cell-based luciferase assay in which AICD generation correlates with increased luciferase expression and induced luminescence emission, did not result in a significant inhibition of AICD release (Supplementary Fig. 1b). Additionally, the GSMs were tested in human lymphocytic leukemia SUP-T1 cells constitutively producing NICD. In contrast to the GSI XIX which completely abolished NICD production, E2012 and GSM-1 did not affect NICD production at concentrations up to 10 mM (Supplementary Fig. 2a). These findings were confirmed in a cellular Notch-based assay in which a luciferase reporter system is directly placed under the control of NICD²⁴ (Supplementary Fig. 2b), as well as in HEK cells overexpressing NotchDE²⁵ (Supplementary Fig. 2c). In the latter two assays, E2012 and GSM-1 concentrations up to 1 mM did not affect NICD production. At 10 mM, E2012, but not GSM-1, significantly reduced NICD production in DL4-Hela/N1-Hela co-cultures by 55% (**p<0.001) (Supplementary Fig. 2b). Taken together, we conclude that E2012 and GSM-1 do not alter the Notch and APP ϵ -site cleavage at concentrations as high as 10x cellular IC₅₀. At higher concentrations (100x cellular IC₅₀), only the non-NSAID-based compound E2012 can inhibit to some extent intracellular signaling of Notch and APP in cell-based assays.

APP mutations causing familial Alzheimer's disease differentially alter g- and ϵ -sites.

APP mutations causing familial Alzheimer's disease (FAD) can be grouped based on their localization at three different regions: at the b-secretase cleavage site, in close vicinity of the a-secretase cleavage site in the A β region, or in the transmembrane domain (TMD) of APP, between the g- and the ϵ -sites (Fig. 2a, left). These mutations differ not only by the onset of the disease (Fig. 2a, right), but also by the molecular fingerprint of the produced A β peptides. Indeed, the vast majority of mutations in the a- and b-sites lead to overproduction of total A β (mutations in the A β region can also render the A β 42 peptide more fibrinogenic and prone to aggregation) while mutations in the transmembrane region seem to be linked to an elevated A β 42/40 ratio^{12,26}. Because no information exists about how these mutations affect both the g- and ϵ -cleavage sites in APP, and because AICD has been suspected to play a role in the pathogenesis of Alzheimer's disease, we analyzed A β and AICD products generated from purified APP-C99 substrates carrying the FAD mutations listed in Figure 2a. All recombinant substrates were processed by purified γ -secretase (Supplementary Fig. 3). IP/MS analyses of the A β and AICD profiles revealed that amino acid substitutions clustered near the a-secretase (A692G, E693G, D694N) apparently do not alter the A β or AICD profiles, when compared to the control WT APP-C99 (Fig. 2b). In striking contrast, T714I and V175A mutations close to the transmembrane γ -cleavage site lead to drastically increased A β 42/40 and A β 38/40 ratios and to a total inhibition of the production of the shortest AICD50-99 (Fig. 2c). Interestingly, mutations closer to the ϵ -cleavage sites (I716V, V717F and L723P) are characterized by an intermediate phenotype as they do not alter A β profiles, but clearly lower the production of AICD50-99 (below detection levels for V717F and L723P) (Figs. 2b and c). Taken together, our data demonstrate that FAD-causing mutations in the transmembrane region of APP, but not close to the a-secretase site, affect both g- and/or ϵ -sites.

GSMs affect the g- but not the ϵ -cleavage of APP with FAD mutations.

To better understand the molecular mechanism by which GSMs shift A β production, as well as to verify whether GSMs can still lower A β 42 and have potential therapeutic benefits in patients with FAD, we decided to test GSM-1 on APP-C99 with FAD mutations. GSM-1 was chosen because it was more potent in our *in vitro* γ -secretase assay, and also because this compound did not affect e-cleavage in all cell-based experiments. First, 10 mM GSM-1 did not affect the levels of total cleavage products (compare AICD and A β levels in Supplementary Figs. 3a and 3b), providing further evidence that GSMs do not modify the overall yield of γ -secretase processing. Next, we found that GSM-1 conserved its A β 42-lowering and A β 38-raising properties on all tested mutant substrates (Fig. 3a). However, GSM-1 did not influence the e-cleavage and none of the mutants tested recovered the wild-type phenotype of nearly equimolar production of AICD49-99 and AICD50-99 (compare AICD profiles in Figs. 3b and 2c). Our findings provide new evidence that GSMs lower A β 42 production by interfering with the g-site in APP substrates with FAD mutations, strongly suggesting that none of the substituted amino-acid residues are crucial for the GSM-dependent modulation of APP processing by γ -secretase. They further demonstrate that GSMs fail to reverse the effects on AICD production of FAD-causing mutations in the transmembrane domain of APP.

Structural changes in the TM domain of APP with FAD-causing mutations

To support our results mechanistically, we used molecular dynamics (MD) simulations to further insights into the structural and dynamic properties of the transmembrane domain (TMD) of wild-type APP and three FAD-mutants (T714I, V715A and V717F) in the membrane environment. These mutations were chosen because they either affect both γ - and ϵ -cleavage sites (T714I, V715A), or exclusively the ϵ -site (V717F). The TM region of the APP-C99 NMR structure (PDB code 2LP1²⁷) was used and inserted into a phosphatidylcholine (POPC) lipid bilayer (Fig. 4a). MD simulations were carried out under *vacuo* physiological conditions (i.e., pH ~7, 1 atm, 300 K), mimicking as close as possible the conditions of the γ -secretase enzymatic assay used in our study. For all systems, the secondary structure of APP fluctuated only marginally. Consistent with the NMR observations, the N-terminus amphiphilic helix (Q686–N698) lies on the membrane surface, and a lysine belt (K687, K699, K724–726) anchors the TMD and N-terminus helices, respectively, in and onto the membrane (Fig. 4a). During simulation, the helices tilted with respect to the membrane surface. We found that the V717F mutation affected the most the orientation of the TMD, tilting it by 45 ± 5 degrees, while the wild type and T714I and V715A mutants remained more vertical (36 ± 5 , 38 ± 3 and 36 ± 3 degrees, respectively) (Fig. 4a and Supplementary Fig. 4a). These mutations further shifted the TMD perpendicularly to the membrane surface. When measuring the position of the A β 42 cleavage site (I712 - A713 peptide bond), T714I and V715A mutants shifted closer (2.9 ± 1.2 Å and 2.3 ± 0.9 Å, respectively) to the cytoplasmic interface compared to the WT, whereas V717F moved closer (-0.2 ± 1.3 Å) to the extracellular membrane layer (Fig. 4b). The combined effect of elevation and tilting perturbations of the APP TMD due to FAD-mutations is probably affecting the presentation to the enzyme of the γ -cleavage site at position 42, correlating quantitatively with the changes in A β 42/A β 40 ratio (Fig. 2). Moreover, the flexibility of the transmembrane domain, estimated based on the local bending of the helix²⁸, resulted to be much higher for T714I and V715A mutants, specifically at the G₂₉XXXGXXXG₃₇ region (Fig. 4c). Molecular docking and simulations of GSM-1 compound suggested the same region as the conserved drug-binding pocket (L701, G704 and G708 being the most direct interactions with GSM-1, Supplementary Fig. 4) for both the wild-type and mutant APP. Thus, GSM-1 effect on T714I and V715A FAD mutants might cause an attenuation of their flexibility close to the g-site. For all three mutants, the hydration of the ϵ -site at residue L720 was significantly decreased compared to the WT (Fig. 4d). In particular, for APP-WT one water molecule is in average within 3 Å of L720 backbone, while the water coordination number drops to less than 0.5 Å for V715A, and to nearly zero for T714I and V717F. This effect can be ascribed to the different water permeation properties of the APP C-terminus lysine belt combined with the perturbed helical orientation produced by the mutants. Together, our data support a model in which structural changes and, in particular, the

position of the γ -cleavage site relative to the membrane surface can potentially modulate A β production, while the hydration of L720 can affect the ϵ -cleavage site and the production of the two AICDs potentially implicated in signalling pathways and in the pathogenesis of AD.

DISCUSSION

Mutations in APP cause aggressive early-onset Alzheimer's disease by triggering abnormal production, accumulation and aggregation of A β peptides. The most severe pathological phenotype is observed for mutations occurring in the transmembrane region and in close vicinity of the A β 42 γ -cleavage site, leading to elevated A β 42/40 and A β 38/40 ratios (Fig. 2). Small chemical γ -secretase modulators (GSMs) are known to lower A β 42 and raise A β 38, without affecting the overall production of the A β peptides. In this study, we used second generation GSMs with a strong therapeutic potential (GSM-1 and E2012), in combination with APP substrates with FAD-causing mutations, (i) to better understand the molecular mechanisms of the γ -secretase-dependent, intramembrane and sequential cleavage of the substrate, (ii) to identify changes in APP-cleavage profiles, potentially implicated in the pathogenesis of AD, and (iii) to further support the therapeutic potential of GSMs to safely treat sporadic as well as early-onset familial AD. We further demonstrate here that GSMs represent an attractive scaffold for structure-activity relationship design (SAR) of potent disease-modifying drugs against AD. As they do not affect the processing of the Notch substrate (Fig. 1), GSMs are expected to circumvent clinical side effects observed with GSIs and attributed to impaired Notch processing¹⁷. Moreover GSMs do not alter AICD production (Fig. 1) and thus are not likely to influence the downstream cell signaling functions of APP, again underlying the huge therapeutic potential of these compounds in sporadic AD cases. Additionally, our results confirm altered g-site processing in APP-FAD mutations, associated with an increase of A β 42 and A β 38 and demonstrate that mutations located in close vicinity to the g-site do impede the production of the shorter ϵ -site counterpart AICD50-99. Mutations T714I and V715A display the strongest phenotype at both g- and ϵ -sites, with highest A β 42/40 and A β 38/40 ratios. Moreover, these mutations lead to the production of only one detectable AICD, suggesting a putative role of AICD in AD pathogenesis. These results might be explained with the sequential cleavage model of APP-CTF, in which the generation of A β 40 is dependent on the AICD50-99, while A β 42/38 is linked to AICD49-99¹⁴. Reduced AICD50-99 (as observed for some FAD mutants) is expected to be associated with less A β 40 and more A β 42/38, thus increasing the pathogenic ratio of A β 42/40. The combination of higher A β 42/40 ratio and the loss of AICD50-99 can explain the extreme aggressiveness of these APP mutations with regard to the onset of the disease. GSMs were still potent at the g-site of mutated APP substrates, but failed to rescue the phenotype at the ϵ -cleavage site (Fig. 3). These findings clearly demonstrate that GSMs can be used for the treatment of patients with both sporadic and genetic forms of AD. To better understand the effect of GSM binding, we docked the GSM-1 compound on the ensemble of wild-type and FAD-mutants (T714I, V715A and V717F) conformations generated by MD simulations. GSM-1 strongly binds to the previously identified G₂₉XXXGXXXG₃₇ motif (Supplementary Fig.4), further supporting our observation that amino acids close to the g-sites of APP are not required for GSMs activity, again underlying the therapeutic potential of these compounds²⁰.

Rationalizing these results with structural models based on molecular simulations of wild type and mutant APP substrates, we propose that the observed effects on A β and AICD production can be triggered by combined changes in the flexibility, tilting angle, and positioning of both g- and ϵ -sites in the membrane, along with the hydration levels at the ϵ -sites. Such structural changes, by potentially modifying the presentation of the substrates to the enzymatic complex, might affect the γ -secretase cleavage specificity. Consistent with this model, the elevated A β 38/40 ratio observed in mutants T714I and V715A can be explained by the vertical shift of the helix in the membrane bilayer, with the consequent presentation of position 38 instead of 40 to the γ -secretase cleavage site. The observed vertical shift in mutants (~2.5 Å) is in fact associated with a tilting of half of a helix turn (i.e., displacement of 2 amino acids). Interestingly,

A β 38 and A β 42 have been proposed to originate from the same sequential processing pathway, starting at the e-position 48, and directly linked to AICD49-99 production¹⁴. Consistent with this sequential model, we propose that the conformational changes in the T714I and V715A substrates, associated with reduced AICD50-99 levels, explain the preferential A β 38 and A β 42 production observed with these mutants. Interestingly, and apparently in contrast to the sequential cleavage model, the V717F and L723P mutants, although characterized by reduced AICD50-99 levels, do not show modifications in their A β profiles. This observation may reflect the existence of two or more different pathological molecular explanations, classifying FAD forms in different categories.

The nearly complete loss of AICD50-99 production in mutants T714I, V715A, V717F and L723P, can be attributed, according to our model, to higher solvation states around positions 49 and 50, with the latter becoming more exposed to water molecules. One can speculate that the increased water accessibility to the leucine 49 may favor the cleavage at this position, or inhibit the cleavage at valine 50 by preventing the interaction at position 50 of APP-C99 with the hydrophobic transmembrane domain 1 of the γ -secretase catalytic component PS1.

Together, our data reported here provide new insights into the therapeutic mechanism and potential of chemical γ -secretase modulators. They are further expected to improve our current understanding of the pathological mechanism of familial AD.

METHODS

γ -secretase inhibitors and modulators. γ -secretase inhibitors Compound E and GSI XIX were purchased from Merck Millipore (Darmstadt, Germany). The heteroaryl-type GSM E2012 (US2006/0004013) was obtained from Haoyuan Chemexpress Co. and the NSAID-type GSM-1 was synthesized at Syngene according to methods described in WO2006/043064.

Molecular Biology.

The pET21b vector encoding the cDNAs of human APP-C99-His was used as a template for the generation by PCR-amplification of all FAD APP-C99 mutants described in this study. The constructs were expressed in BL21(DE3) cells induced with 1 mM isopropyl- β -D-thiogalactopyranoside (IPTG) for 4 h at optic density 600 nm = 0.8. For the purification of APP-C99 mutants, harvested bacteria were lysed with 10 mM Tris pH 7.0, 150 mM NaCl, 1 % Triton X-100, protease inhibitor cocktail (PI; Roche, Rotkreuz, Switzerland), and passed through a high-pressure homogenizer (Emulsiflex-C5; Avestin, Inc., Mannheim, Germany). The obtained lysates were spun down at 5000 x g for 30 min and supernatants were incubated overnight with Ni-NTA agarose beads (Invitrogen, Basel, Switzerland) at 4 °C. Bound proteins were eluted in 1 % NP-40 containing 250 mM imidazole (pH 7.8), analyzed by Coomassie stained sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and quantified by Bicinchoninic Acid (BCA; Pierce, Rockford, IL, USA). The normalization was validated by WB. Recombinant APP-C100-Flag and Notch N100-Flag substrates were also expressed in BL21(DE3) cells, but 100 mM glycine (pH 2.7)/1% NP40 was used for elution from M2-agarose beads.

Western blotting and antibodies.

Samples from γ -secretase activity assays were run on 4–12 % Bis-Tris gels and transferred onto PVDF membranes to detect Ab, Flag-tagged APP- or Notch-intracellular domain (AICD-Flag and NICD-Flag) with Ab-specific 6E10 (1:1000; Covance, Berkeley, CA, USA) and Flag-specific M2 antibodies. Anti-

mouse/rabbit/rat IgG conjugated to Alexa 680 were purchased from Invitrogen and the Odyssey infrared imaging system (LICOR) was used to detect the fluorescent signal.

Whole cell extracts prepared in 50 mM HEPES buffer containing 1 % NP40 were separated by electrophoresis on NuPAGE® Novex® 4-12 % Bis-Tris Gels (Invitrogen, Carlsbad, CA, USA) for SDS-PAGE analysis transferred onto nitrocellulose membranes, and probed with antibodies against APP-CTFs (Sigma-Aldrich GmbH, Buchs, Switzerland.). For detection of NICD, membranes were blocked with 5 % milk and incubated overnight with primary antibody Val1744 anti-Notch (Cell Signaling, Bioconcept, Allschwil, Switzerland) at 4 °C. Membranes were washed with 1x TBST (1x TBS + 0.1 % Tween 20) for 15 min and incubated with HRP-conjugated secondary antibodies for 1 h at room temperature. Signal was detected with Super Signal West chemiluminescent substrate (Thermo Scientific, catalog number 34077).

γ-secretase purification. γ-secretase used in this study was purified from 10 liters of PB suspension cultures by following the protocol described by Alattia et al. (submitted for publication).

γ-secretase activity assays. γ-secretase assays using the recombinant APP- or Notch-based substrates were performed as previously reported²⁹. Briefly, highly-purified γ-secretase were solubilized in 0.2 % (wt/vol) CHAPSO, 50 mM HEPES (pH 7.0), 150 mM NaCl, 5 mM MgCl₂ and 5 mM CaCl₂, and incubated at 37 °C for 4 h with 1 mM substrate, 0.1 % (wt/vol) PC, 0.025 % (wt/vol) PE. The resulting products, APP intracellular domain AICD-Flag, AICD-His, Ab, and Notch intracellular domain NICD-Flag were detected with above-described antibodies, or analyzed by mass spectrometry.

Immunoprecipitation-mass spectrometry (IP-MS) analysis of Aβ and AICD. Aβ and AICD generated in γ-secretase *in vitro* assays with purified substrates were analyzed as previously described²⁹. Briefly, Aβ was IP'd overnight using anti-Aβ antibody 4G8 (Covance, Berkeley, CA, USA) and protein G coupled to agarose resin (Roche Applied Science, Penzberg, Germany). For AICD or NICD detection, 12 mL 10 % Triton-X100 solution was added after the enzymatic reaction and incubated for 20 min at 55 °C prior to overnight immunoprecipitation at 4 °C with M2-anti Flag resin (Invitrogen, Carlsbad, CA, USA), or Protein G (Roche, Mannheim, Germany), coupled with anti-C-terminal AICD antibody. Aβ and AICD-Flag were eluted with 1:20:20 (v:v:v) 1 % (vol/vol) trifluoroacetic acid:acetonitrile:H₂O, equally mixed with saturated CHCA (α-cyano 4-hydroxy cinnaminic acid), and analyzed by MALDI-TOF mass spectrometry in reflectron mode on an ABI 4800 MALDI-TOF/TOF mass spectrometer (Applied Biosystems, Carlsbad, CA, USA). Molecular masses were accurately measured and searched against amino acid sequences of human APP-C99 with addition of a methionine residue at the N-terminus and a Flag tag sequence (C100Flag) or a His tag sequence (C100His) at the C-terminus, or a A692G, E693G, D654N, T714I, V715A, I716V, V717F and L723P substitution.

Molecular dynamics simulations and molecular docking. We used molecular dynamics (MD) simulations to characterize the TM domain of WT and APP with FAD mutations. In a first step, the NMR structure (PDB code 2LP1²⁷) was inserted and equilibrated in a 60 x 60 Å² Palmitoyl Oleoly Phosphatidyl Choline (POPC) patch³⁰ to relax the structure in a phospholipid bilayer. The system was solvated in 60 x 60 x 100 Å² water box, neutralized through the addition of NaCl at a concentration of 150 mM. After 30 ns, the equilibrated structure was used as initial seed to build the other three mutants (V714A, T715I and V717F) and WT-ligand complex. GSM-1 ligand was parameterized using CGenFF and ParamChem 0.9.6³¹. All MD simulations were performed using NAMD 2.8 engine, with the CHARMM27 force field, including CMAP corrections for the protein and CHARMM36 for the POPC membrane^{32,33}. TIP3P water parameterization was used to describe the water molecules. The periodic electrostatic interactions were computed using the particle-mesh Ewald (PME) summation with a grid spacing smaller than 1 Å. All systems were equilibrated for 30 ns at 300K. Free molecular dynamics were performed up to 100 ns with a 2 fs integration time step using the RATTLE algorithm applied to all bonds. Constant temperature (300 K) was imposed by using Langevin dynamics, with damping coefficient of 1.0 ps. Constant pressure of 1 atm was maintained with a Langevin piston dynamics, 200 fs decay period and 50 fs time constant.

GSM-1 was docked on WT and APP mutants using SWISSDock^{34,35} and BSP-SLIM³⁶; independently both approaches produced similar poses. Since these algorithms do not account for the presence of the membrane, we discarded all unphysical docking poses, e.g. the GSM-1 hydroxyl being buried in the hydrophobic membrane core. The conformation with the lowest energy was then selected for each system.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation (to P.C.F., grant 31003A_134938/1, to M.D.P. grant 200020_138013), the Swiss National Centres of Competence in Research, Neural Plasticity and Repair (to P.C.F.), the Strauss foundation (to P.C.F.), and an EPFL-Merck-Serono grant (to J.R.A. and M.D.). The authors thank M. Moniatte (Proteomics Core Facility, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland) for technical assistance with mass spectrometry, Laurence Anderes for the Notch assay in SUP-T1 cells, and M.S. Wolfe for the T20 cell line. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

M.D., L.R., J.R.A., J.H and A.F. performed experiments; T.L. performed and analyzed with M.D.P. molecular simulations; I.H., F.R., D.B., M.D.P. and P.C.F. designed and supervised the study; M.D. and P.C.F. wrote the manuscript. All authors edited the paper.

COMPETING FINANCIAL INTERESTS

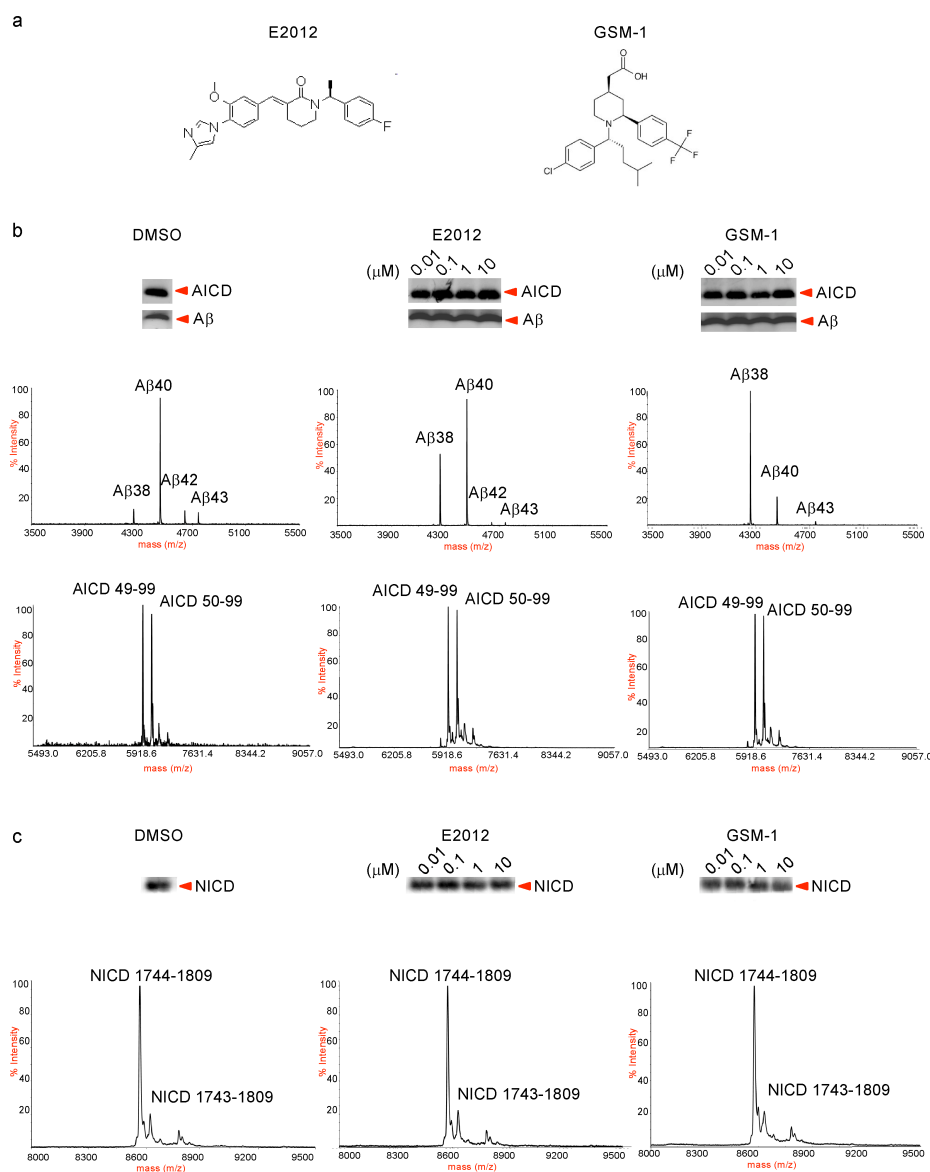
D.B. and I.H. are full time employees of Merck Serono S.A. Other authors declare having no competing financial interests.

REFERENCES

1. Masters, C.L. et al. Amyloid plaque core protein in Alzheimer disease and Down syndrome. *Proc Natl Acad Sci U S A* **82**, 4245-9 (1985).
2. Robakis, N.K. et al. Chromosome 21q21 sublocalisation of gene encoding beta-amyloid peptide in cerebral vessels and neuritic (senile) plaques of people with Alzheimer disease and Down syndrome. *Lancet* **1**, 384-5 (1987).
3. Goate, A.M. et al. Predisposing locus for Alzheimer's disease on chromosome 21. *Lancet* **1**, 352-5 (1989).
4. Sherrington, R. et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* **375**, 754-60 (1995).
5. Haass, C. et al. Normal cellular processing of the beta-amyloid precursor protein results in the secretion of the amyloid beta peptide and related molecules. *Ann N Y Acad Sci* **695**, 109-16 (1993).
6. De Strooper, B. et al. A presenilin-1-dependent gamma-secretase-like protease mediates release of Notch intracellular domain. *Nature* **398**, 518-22 (1999).
7. Bot, N., Schweizer, C., Ben Halima, S. & Fraering, P.C. Processing of the synaptic cell adhesion molecule neurexin-3beta by Alzheimer disease alpha- and gamma-secretases. *J Biol Chem* **286**, 2762-73 (2011).
8. Uemura, K. et al. Characterization of sequential N-cadherin cleavage by ADAM10 and PS1. *Neuroscience Letters* **402**, 278-283 (2006).
9. Ohki, Y. et al. Phenylpiperidine-type gamma-secretase modulators target the transmembrane domain 1 of presenilin 1. *EMBO J* **30**, 4815-24 (2011).
10. Lazarov, V.K. et al. Electron microscopic structure of purified, active gamma-secretase reveals an aqueous intramembrane chamber and two pores. *Proc Natl Acad Sci U S A* **103**, 6889-94 (2006).
11. Cacquevel, M. et al. Rapid purification of active gamma-secretase, an intramembrane protease implicated in Alzheimer's disease. *J Neurochem* **104**, 210-20 (2008).
12. Chavez-Gutierrez, L. et al. The mechanism of gamma-Secretase dysfunction in familial Alzheimer disease. *EMBO J* **31**, 2261-74 (2012).
13. Haass, C. et al. The Swedish mutation causes early-onset Alzheimer's disease by [beta]-secretase cleavage within the secretory pathway. *Nat Med* **1**, 1291-1296 (1995).
14. Sato, T. et al. Potential link between amyloid beta-protein 42 and C-terminal fragment gamma 49-99 of beta-amyloid precursor protein. *J Biol Chem* **278**, 24294-301 (2003).
15. Eckman, C.B. et al. A new pathogenic mutation in the APP gene (I716V) increases the relative proportion of A beta 42(43). *Hum Mol Genet* **6**, 2087-9 (1997).
16. Kumar-Singh, S. et al. Nonfibrillar diffuse amyloid deposition due to a gamma(42)-secretase site mutation points to an essential role for N-truncated A beta(42) in Alzheimer's disease. *Hum Mol Genet* **9**, 2589-98 (2000).
17. Schor, N.F. What the halted phase III γ -secretase inhibitor trial may (or may not) be telling us. *Annals of Neurology* **69**, 237-239 (2011).
18. Weggen, S. et al. A subset of NSAIDs lower amyloidogenic Abeta42 independently of cyclooxygenase activity. *Nature* **414**, 212-6 (2001).
19. Eriksen, J.L. et al. NSAIDs and enantiomers of flurbiprofen target gamma-secretase and lower Abeta 42 in vivo. *J Clin Invest* **112**, 440-9 (2003).
20. Kukar, T.L. et al. Substrate-targeting gamma-secretase modulators. *Nature* **453**, 925-9 (2008).

21. Crump, C.J. et al. Piperidine acetic acid based gamma-secretase modulators directly bind to Presenilin-1. *ACS Chem Neurosci* **2**, 705-710 (2011).
22. Mangialasche, F., Solomon, A., Winblad, B., Mecocci, P. & Kivipelto, M. Alzheimer's disease: clinical trials and drug development. *Lancet Neurol* **9**, 702-16 (2010).
23. Bulic, B. et al. Chemical Biology, Molecular Mechanism and Clinical Perspective of gamma-Secretase Modulators in Alzheimer's Disease. *Curr Neuropharmacol* **9**, 598-622 (2011).
24. O'Neil, J. et al. FBW7 mutations in leukemic cells mediate NOTCH pathway activation and resistance to gamma-secretase inhibitors. *J Exp Med* **204**, 1813-24 (2007).
25. Liao, Y.-F., Wang, B.-J., Cheng, H.-T., Kuo, L.-H. & Wolfe, M.S. Tumor Necrosis Factor-alpha, Interleukin-1beta, and Interferon-gamma Stimulate gamma-Secretase-mediated Cleavage of Amyloid Precursor Protein through a JNK-dependent MAPK Pathway. *Journal of Biological Chemistry* **279**, 49523-49532 (2004).
26. Walsh, D.M., Hartley, D.M., Condron, M.M., Selkoe, D.J. & Teplow, D.B. In vitro studies of amyloid beta-protein fibril assembly and toxicity provide clues to the aetiology of Flemish variant (Ala692-->Gly) Alzheimer's disease. *Biochem J* **355**, 869-77 (2001).
27. Barrett, P.J. et al. The amyloid precursor protein has a flexible transmembrane domain and binds cholesterol. *Science* **336**, 1168-71 (2012).
28. Dahl, A.C., Chavent, M. & Sansom, M.S. Bendix: intuitive helix geometry analysis and abstraction. *Bioinformatics* **28**, 2193-4 (2012).
29. Alattia, J.R. et al. Mercury is a direct and potent gamma-secretase inhibitor affecting Notch processing and development in Drosophila. *FASEB J* **25**, 2287-95 (2011).
30. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-8, 27-8 (1996).
31. Vanommeslaeghe, K. et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* **31**, 671-90 (2010).
32. Phillips, J.C. et al. Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-802 (2005).
33. Klauda, J.B. et al. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J Phys Chem B* **114**, 7830-43 (2010).
34. Grosdidier, A., Zoete, V. & Michielin, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res* **39**, W270-7 (2011).
35. Grosdidier, A., Zoete, V. & Michielin, O. Fast docking using the CHARMM force field with EADock DSS. *J Comput Chem* (2011).
36. Lee, H.S. & Zhang, Y. BSP-SLIM: a blind low-resolution ligand-protein docking approach using predicted protein structures. *Proteins* **80**, 93-110 (2012).
37. Jeffries, S., Robbins, D.J. & Capobianco, A.J. Characterization of a high-molecular-weight Notch complex in the nucleus of Notch(ic)-transformed RKE cells and in a human T-cell leukemia cell line. *Mol Cell Biol* **22**, 3927-41 (2002).

FIGURE LEGENDS



Figures 1. GSMs alter g-site processing, but do not modulate ϵ -site on APP and Notch *in vitro*. **a)** Chemical structure of clinically tested small molecules NSAID-based phenylpiperidine-type GSM-1 and bridged aromatics E2012. **b)** GSMs do not inhibit the overall cell-free processing of APP as shown by WB analysis of Aβ and AICD at concentrations ranging from low nanomolar to low micromolar. Following immunoprecipitation, analysis by MALDI-TOF demonstrated the efficacy of GSMs on the g-site, where they shifted the production towards shorter Aβ38 and lowered Aβ42. Even at concentrations as high as 10 mM, GSMs do not alter the ϵ -site as evidenced by the IP/MS on the AICD products, AICD49-99 and AICD50-99. **c)** GSMs show no inhibition on Notch intracellular signaling products. *In vitro* NICD formation detected by WB is unaltered. The ratio of the two NICD products, NICD1744-1809 and NICD1743-1809, detected after IP/MS, also remained without change in the presence of 10 mM of both compounds.

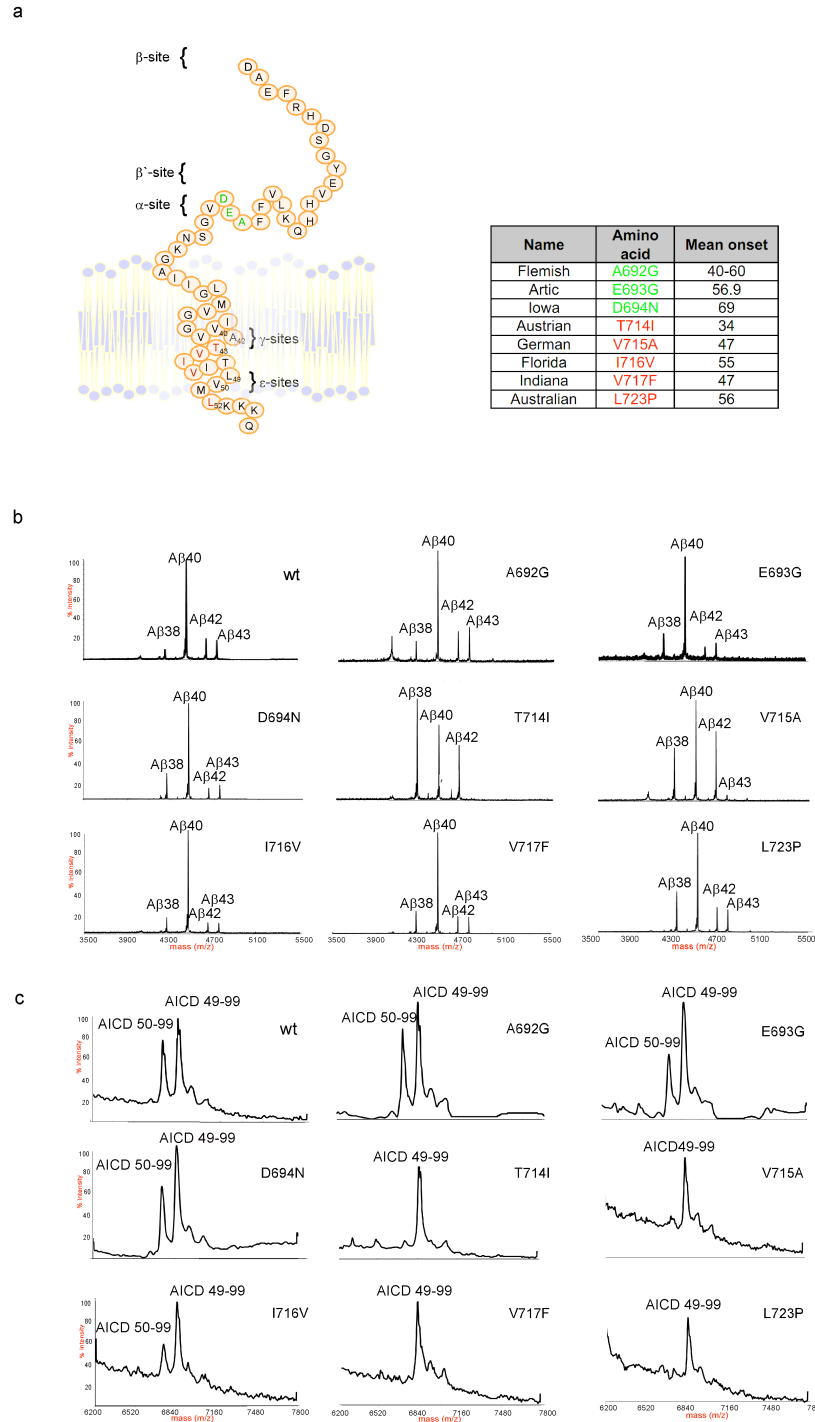


Figure 2. Pathological mutations in APP cause altered processing at both e- and g-sites. **a)** FAD-APP mutations differ by the onset, but are all clustered close to the α-secretase and γ-secretase cleavage sites, as represented schematically in the left panel (color coding: α-secretase site mutations in green; γ-secretase site mutations in red). Different mean onsets are summarized in the right table. The earliest onset is reported for Austrian and German families at the age of early 50s. **b)** IP/MS profiles of Aβ

generated from purified FAD-APP-C99 substrates show that mutations A692G, E693G, D694N, I716V, V717F and L723P have similar and WT-like A β 38/40 and A β 42/40 ratios. In contrast, T714I and V715A mutations in the transmembrane region led to a drastic increase in both A β 38 and A β 42 and A β 38/40 and A β 42/40 ratios. **c)** AICD profiles from the same reactions as analyzed by IP/MS. FAD-APP-C99 mutations close to the a-secretase site produce AICD49-99 and AICD50-99 in nearly equimolar ratios. Amino acid substitutions in close vicinity of the g-sites in the TMD strongly affect the yield of shorter AICD50-99.

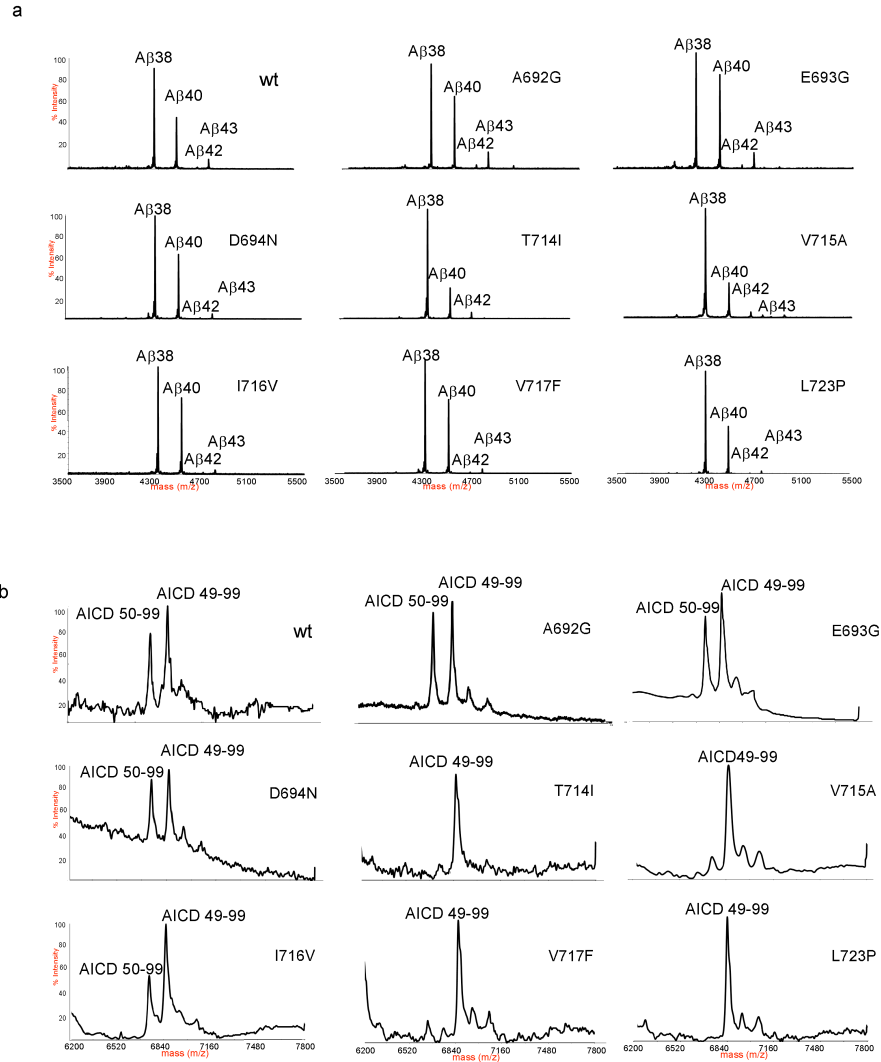


Figure 3. GSMs potentially shift the production at g-, but not at ϵ -sites in APP with FAD mutations. **a)** IP/MS profiles of A β species produced from purified APP substrates with FAD mutations detected by IP/MS demonstrate the efficacy of GSMs towards the g-site processing. For all FAD-APP mutants tested, A β 38/40 and A β 42/40 ratios are strongly affected by 10 mM of GSM-1. Mutations T714I and V715A had the most elevated A β 38/40, implying that amino acids between APP g- and ϵ -sites are not essential for the GSM mechanism of action. **b)** IP/MS profiles of the ϵ -site products from the same reactions provide supplementary evidence that GSMs do not influence the ϵ -sites in APP. Even high concentrations of GSM-1 fail to restore the loss of shorter AICD50-99 in mutants T714I, V715A, V717F and L723P.

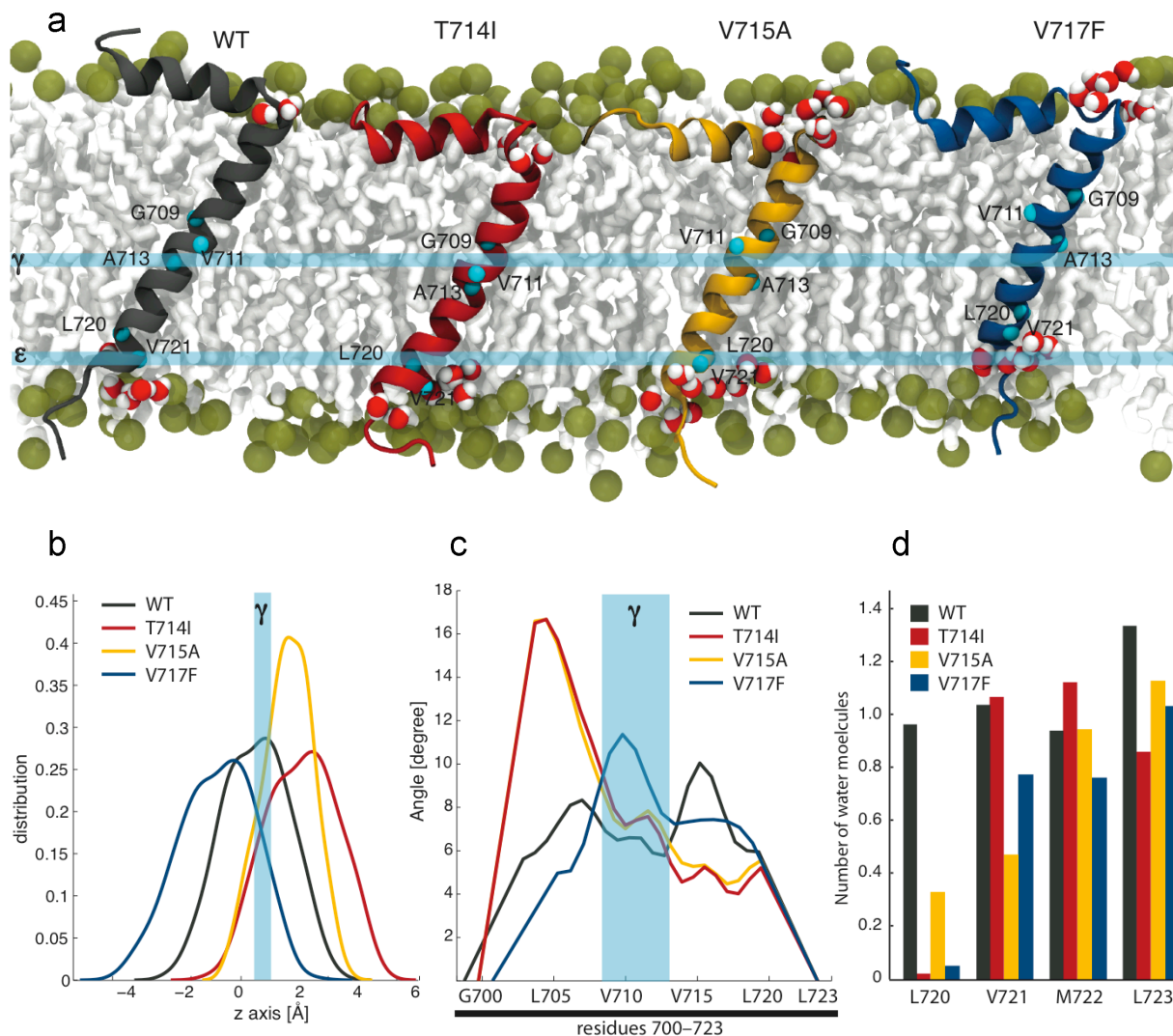


Figure 4. Structural and dynamic models of the TM domain in WT and FAD-causing APP mutants. **a)** The NMR structure of WT APP (PDB code: 2LP1²⁷) and three FAD-causing mutants, namely T714I, V715A and V717F, were inserted in a POPC membrane bilayer and simulated with molecular dynamics. The g-sites G709, V711 and A713 and ε-sites L720 and V721 positions are indicated by blue spheres. **b)** The spatial distribution of Aβ42 cleavage site is reported for the four systems. The distribution is calculated along the axis orthogonal to the membrane bilayer surface and the reference putative position of the g-site in wild-type conditions is highlighted with the blue area. **c)** The local bending of the APP helix is computed for the TM domain (with a standard error of 2.8 degrees). **d)** The coordination number of water molecules around the relevant residues at the ε-site is reported for the WT and mutants.

specific CMG2 gene mutations ((Deuquet et al, 2011a; Deuquet et al, 2011b).

The responsible gene *cmg2* is expressed in all tissues except the brain, consistent with development of normal intelligence in patients (Deuquet et al, 2011b; Stucki et al, 2001). *Cmg2* encodes a 55kDa type I transmembrane protein, consisting of an extracellular N-terminal von Willebrand factor type A domain (vWA) followed by an Ig-like domain, a single transmembrane helix and a cytosolic tail (Deuquet et al, 2011a; Deuquet et al, 2011b). vWA domains, also called I-domain of integrins, contain a metal ion-dependent adhesion site (MIDAS) motif with which they interact with extracellular matrix (ECM) proteins such as collagens (Engel & Chiquet, 2011). The CMG2 vWA domain has been proposed to bind collagen IV and laminin (Bell et al, 2001). The exact function of the CMG2 Ig-like domain has not been determined. It was however shown to contain two disulfide bonds essential for proper CMG2 folding in the endoplasmic reticulum (ER) (Deuquet et al, 2011a). The CMG2 cytoplasmic tail is composed of 148 residues and contains multiple post-translational modification sites such as palmitoylation (Abrami et al, 2006), phosphorylation (Abrami et al, 2010) and ubiquitination (Abrami et al, 2010), modifications that are important for the surface dynamics and plasma membrane turn over of the protein (Deuquet et al, 2011b).

So far, more than 150 cases of HFS have been reported and some 34 different mutations, mostly in exons, have been identified (Deuquet et al, 2011b). Among them, exon 13 is a hot spot for frameshift mutations, which include insertion of one or two bases –c.1073-1074insC (hereafter referred to as insC), c.1073-1074insCC (insCC)– and deletion of one base –c.1074delT (delT). These three frameshift mutations account for approximately 60% of all the pathogenic alleles. The frequency of insertions and deletions at positions 1073-1074 is likely due to its proximity to a low complexity, GC-rich, region encoding a stretch of proline residues (Fig 1A) that could constitute a vulnerable site for errors during DNA replication.

Here, we have investigated the molecular consequences of the frameshift mutations in exon 13, which encodes the beginning of the cytosolic tail of CMG2. We first show that in fibroblasts derived from patients carrying such mutations, the mRNA level of *cmg2* was drastically lower than in fibroblasts from healthy donors. This is most likely due to the generation of a premature stop codon that would be recognized by the nonsense-mediated mRNA decay (NMD) pathway (Rebbapragada & Lykke- Andersen, 2009). This observation would point to the NMD pathway as a potential therapeutic target. We however first studied the effects of the insC, insCC and delT mutations at the protein level by ectopic expression of the corresponding cDNAs in HeLa cells. In this system, we found the insCC and delT mutations lead to mutant CMG2 proteins – CMG2insCC and CMG2^{delT} – that are rapidly targeted to ER associated degradation (ERAD). Patients with this mutation would therefore not benefit from drugs targeting the NMD pathway and consistently silencing *upf1*, a central factor of this mRNA decay pathway, did not rescue the CMG2^{insCC} and CMG2^{delT} protein expression. In contrast, CMG2^{insC}, a mutation that occurs at the same hotspot but causes a different shift in the reading frame, was well expressed in HeLa cells and properly targeted to the cell surface, where it was competent for ligand binding. Importantly, in patient fibroblasts, silencing *upf1* led to a partial restoration of CMG2^{insC} protein, to its proper targeting to the cell surface where it exhibited the ability to bind a surrogate ligand.

This study shows that certain HFS patients could benefit from therapies that would target the NMD pathway. They however also highlights the importance of an in depth analysis of the consequences of specific mutations to determine the appropriate potential treatment.

RESULTS AND DISCUSSION

Analysis of frameshift mutations in HFS patients

The three frameshift mutations in exon 13 of *cmg2* that have been identified in HFS patients all lead to premature stop codons. The mutated insC, insCC and delT genes encode for CMG2 proteins that have modified and shorter cytosolic tails of 29, 67, 66 amino acids respectively, compared to the 148 residue wild type CMG2 tail (Fig. 1A.) Note that the insCC and delT mutations lead to the same altered reading frame and thus to the same cytosolic tail, with the exception of one additional residue at position 359 of CMG2insCC.

To understand the consequences of these frameshift mutations, we analyzed fibroblasts from five previously described patients (Table 1) (Deuquet et al, 2011a; Dowling et al, 2003). Patient 1 is homozygous for the insC single nucleotide insertion (Deuquet et al, 2011a). All other patients are compound heterozygotes, carrying one of the three frameshift mutations on one allele and one missense mutation on the other. Patient 2 (Dowling et al, 2003) carries the insC mutation and a c.566T>C transversion, which causes a p.I198T amino acid substitution (Table 1). Patient 3 (Deuquet et al, 2011a) carries the insC mutation and a c.928G>T transversion resulting in a p.V310F missense mutation (Table 1). Patient 4 (Hanks et al, 2003) carries the insCC mutation and a missense mutation c.566T>C as Patient 2. Patient 5 (Deuquet et al, 2011a) carries the delT mutation and a c.115G>T transversion leading to the p.C39F amino acid substitution (Table 1). All patients suffer from the severe infantile form of HFS, with the exception of patient 3, who suffers from a somewhat milder form possibly due to better pediatric care (Deuquet et al,

2011a; Dowling et al, 2003). We performed a real-time quantitative PCR analysis of mRNA extracted from the patient fibroblasts. The *cmg2* mRNA levels were significantly lower in patient cells than the healthy control fibroblasts (Fig. 1B). Patient 1, the only one carry a frameshift mutation on both alleles, showed the lowest mRNA level. The compound heterozygous patients had *cmg2* mRNA levels ranging from 46% to 72% of that of the control fibroblasts. The decrease of mRNA level observed in patient-derived fibroblasts is presumably due to the non-sense mediated mRNA decay (NMD) pathway, which can recognize the presence of a premature stop codon in mRNA caused by frameshift mutations (Rebbapragada & Lykke-Andersen, 2009). The NMD pathway would thus target both alleles in Patient 1 and only one allele in all other patients.

We next analyzed protein extracts from the patient cells by Western blotting using an anti-human CMG2 monoclonal antibody (Deuquet et al, 2011a). As expected, a band corresponding to full-length CMG2 (55kDa) was readily detected for control fibroblasts. A weaker band of similar molecular weight was detected for cells from the compound heterozygous patients as previously observed (Deuquet et al, 2011a). However none of the truncated CMG2 forms (CMG2^{insC}: 42kDa, CMG2^{insCC} and CMG2^{delT}: 45kDa) could be detected, even at high exposure or following immunoprecipitation using an antibody against the N-terminal ectodomain of CMG2 (data not shown) consistent with the low mRNA levels.

Ectopic expression of CMG2^{insC}, CMG2^{insCC} and CMG2^{delT}

The above observations show that the mRNA levels of the frame-shifted *cmg2* variants are very low, which could account for the absence of CMG2, and suggest that the NMD pathway might be a potential therapeutic target. Before exploring this option, we however first investigated whether the CMG2^{insC}, CMG2^{insCC} and CMG2^{delT} proteins can be properly expressed in tissue culture cells upon transfection with the appropriate cDNAs. Unexpectedly, CMG2^{insCC} and CMG2^{delT} were essentially undetectable by Western blotting whereas CMG2^{insC}, which migrated at an expected molecular weight of 42 kDa, was expressed at similar levels as CMG2^{WT} (Fig 2A). The low expression of CMG2^{insCC} and CMG2^{delT} was not due to reduced synthesis as shown by performing a 5-minute metabolic pulse (Fig. 2B). Immunofluorescence on non-permeabilized cells using a monoclonal antibody directed towards the ectodomain of CMG2 showed that ectopically expressed CMG2^{insC} was properly targeted to the cell surface (Fig. 2C). To investigate whether it was capable of binding ligands, we monitored binding of the anthrax protective antigen (PA), used as a surrogate ligand (Deuquet et al, 2011b; van der Goot & Young, 2009). For this, we used HeLa cells in which the expression of the second anthrax toxin receptor, TEM8, was silenced by lentiviral mediated silencing using an shRNA construct against TEM8 (Fig. S1). As shown both by Western blotting (Fig. 2D) and by immunofluorescence (Fig. 2C), CMG2^{insC} was fully PA-binding competent. Thus CMG2^{insC} retains the ability of WT CMG2 to bind a specific ligand,

in agreement with a previous report (Liu & Leppla, 2003). Since CMG2^{insC} lack most of the WT tail, it is unlikely to retain the full CMG2 function. This is supported by the fact that point mutations in the cytosolic tail have been reported in HFS patients (Deuquet et al, 2011b). The precise consequences of the later mutations have not yet been addressed. Patients homozygous for such mutations however suffer from the mildest form of the disease as opposed to patients carrying the c.1073-1074insC mutation (Deuquet et al, 2011b). Altogether these observations suggest that if the *cmg2* mRNA could be stabilized in patients carrying the c.1073-1074insC mutation, symptoms might be attenuated.

CMG2^{insCC} and CMG2^{delT} are powerful ERAD substrates

The absence of detectable CMG2^{insCC/delT} upon ectopic expression, despite normal synthesis, indicates that these two CMG2 mutants undergo accelerated degradation. A metabolic pulse-chase analysis revealed that indeed the two mutant CMG2 proteins had a half-life of about 20 mins (Fig. 3A). Degradation of CMG2^{insCC} and CMG2^{delT} could be inhibited by treating cells with the proteasome inhibitor MG132 (Fig. 3BC) (Kisselev & Goldberg, 2001). Not only were the levels of metabolically labeled proteins after a 25 min pulse similar to that of CMG2^{WT} (Fig. 3BC), but the proteins could also be detected by western blotting following 4 or 8h MG132 treatment (Fig.3D). In contrast, no rescue of CMG2^{insCC} and CMG2^{delT} was observed following treatment with either Bafilomycin A, an inhibitor of the vacuolar ATPase (Drose & Altendorf, 1997) or leupeptin, an inhibitor of lysosomal proteases (Aoyagi et al, 1969). These observations indicate that CMG2^{insCC} and CMG2^{delT} are rapidly degraded by the proteasome and not in lysosomes. We next investigated whether the MG132 rescued proteins were properly trafficked through the biosynthetic pathway. We have previously shown that CMG2 is glycosylated (Abrami et al, 2006; Deuquet et al, 2009). Due to the modifications of the N-linked glycans in the Golgi, CMG2 migrates as a smear that is insensitive to treatment with Endoglycosidase H (Endo H), an enzyme that only removes the non-complex oligosaccharides. While the majority of expressed CMG2^{WT} and CMG2^{insC} were insensitive to EndoH treatment, MG132 rescued CMG2^{insCC} and CMG2^{delT}, which migrated as a well-defined band, was EndoH sensitive (Fig. 3E), indicating that these mutant CMG2 molecules were unable to reach the Golgi. Altogether,

the above observations show that CMG2^{insCC} and CMG2^{delT} are rapidly degraded by the ER associated degradation (ERAD) pathway and that even MG132-rescued CMG2^{insCC} and CMG2^{delT} cannot exit the compartment. This is in marked contrast to previously characterized HFS CMG2 mutants for which surface expression could be rescued with proteasome inhibitors (Deuquet et al, 2011a). Consistent with the notion that CMG2^{insCC} and CMG2^{delT} are efficient ERAD substrates, we found that the MG132 rescued proteins were polyubiquitinated (Fig. 3F), a hallmark of proteins destined to degradation. This ubiquitination signature was markedly different to the mono- ubiquitination of CMG2 observed during anthrax toxin endocytosis (Fig. 3F, cells treated with PA) (Abrami et al, 2010).

Structural predictions of the CMG2^{insCC/delT} tail

In order to understand why CMG2^{insCC/delT} are powerful ERAD substrates, we analyzed the primary sequence of their cytosolic tail (Fig. 1A) and compared it to that of the WT protein. Little is known about the structure of the cytosolic tail of CMG2^{WT}. We tested five different secondary structure prediction algorithms which all indicated that the CMG2^{WT} tail is mostly composed of random coil (Fig. 4A) as is the much shorter CMG2^{insC} tail. We therefore performed disorder predictions (Dosztanyi et al, 2005), which indeed predicted that residues 342 to \approx 450 are highly disordered (Fig. 4B, left panel). Disordered domains generally are rich in charged residues and poor in hydrophobic residues (Dyson & Wright, 2005; Tompa, 2012) and consistently, the calculated hydrophobic index (Gasteiger E., 2005) was negative along the entire length of the cytoplasmic tail of CMG2^{WT} (Fig. 4C), as it was for CMG2^{insC}. Similar analyses on the CMG2^{delT} tail led to markedly different results. The CMG2^{delT} tail was predicted to be largely structured, based both on secondary structure (Fig. 4A) and disorder (Fig. 4B, right panel) predictions. Moreover most of the tail showed a positive hydrophobic index (Fig. 4C, right panel). To further strengthen these predictions, we investigated the structure of the CMG2^{delT} tail using molecular dynamics (MD) under physiological conditions. The CMG2 helical transmembrane domain was inserted into a lipid membrane bilayer and linked to the CMG2^{delT} cytosolic tail, initially modeled as random coil. During the MD simulation, the tail spontaneously folded into a domain showing three structured α -helices, in good agreement with the secondary structure prediction for the CMG2^{delT} tail (Fig. 4B). The hydrophobic core produced by the helical arrangement stabilizes the cytosolic domain in a globular fold (Figs. 4C, D), which remains in close contact with the membrane surface due to the presence of several lysine residues (K361, K365, K399 and K404) interacting with the lipid head-groups. Thus the insCC and delT hotspot mutations lead to the synthesis of a cytosolic tail that has drastically different properties from that of the original CMG2 tail. Instead of being intrinsically disordered and hydrophilic, the CMG2^{insCC/delT} tail is predicted to form a compact well-folded domain. This domain however exposed hydrophobic patches that are likely to be recognized by the ER quality control machinery and thus targeted to ERAD (Hegde & Ploegh, 2010). These findings are rather counterintuitive in the sense that the native CMG2 tail despite its lack of predicted structure is not seen by the ER quality control system as problematic. In contrast the CMG2^{insCC/delT} tail, although well structured, is immediately spotted.

The CMG2^{insCC/delT} tail sequence is sufficient for targeting to degradation

In order to experimentally validate some of the above predictions, we performed Circular Dichroism (CD) on a synthetic peptide corresponding to the sequence of the CMG2^{delT} tail. The far UV CD spectrum of the peptide was typical of that of polypeptide rich in α -helices, with two negative peaks at 222 and 208nm (Fig 5A). We next analyzed the capacity of the peptide to bind 8-anilino-1-naphthalene-sulfonic acid (ANS), a fluorescent dye that increases its quantum yield when binding hydrophobic patches (Slavik, 1982). As a control peptide, we used a 40 residues peptide of the human β -amyloid (Iacovache et al, 2011). Unfortunately, we were unable to produce the WT CMG2 cytosolic tail recombinantly, presumably due to its unstructured nature. As shown in Fig. 5B, the CMG2^{delT} tail peptide showed strong ANS binding capacity, confirming the predicted presence of surface-exposed hydrophobic patches. We next analyzed whether the CMG2^{delT} tail sequence is sufficient to target a transmembrane protein to ERAD. We therefore generated a chimera between CD4 and the CMG2^{delT} tail. CD4 is a type I transmembrane protein widely used in the construction of chimeric proteins to study membrane trafficking and protein targeting (Jackson et al, 1990; Nilsson et al, 1989; Shin et al, 1991). It has a very stable ectodomain containing 3 disulfide bonds, a single transmembrane helix and a 40 residue cytoplasmic tail. This tail was swapped with that of CMG2^{delT} generating CD4-tail^{mut}. While CD4 expression was readily detectable upon ectopic expression in Hela cells, CD4-tail^{mut} expression was very low (Fig. 5C, top panel), even though synthesis was normal (Fig. 5C, bottom panel). This was due to post-translational degradation via the proteasome, as revealed by the ability of MG132 to rescue expression (Fig. 5C, top panel). We also fused the CMG2^{delT} tail sequence to a soluble cytosolic protein, namely EGFP (EGFP-tail^{mut}), and found that it was also rapidly targeted to degradation via the proteasome (Fig. 5D). Thus the CMG2^{delT} tail sequence is a well-folded domain, which exposes hydrophobic patches. Most likely due to this exposed hydrophobicity, the domain is efficiently recognized by quality control systems both at the surface of the ER and in the cytosol, targeting the protein to proteasomal degradation. These observations show that targeting the

NMD pathway in patients carrying the insCC or delT frameshift mutations is pointless, since the mRNA that would be rescued encodes for a protein that is immediately be recognized by the folding quality control machinery and targeted to degradation at the ER level.

Inhibiting NMD pathway can rescue CMG2^{insC}

Figure 2 illustrates that, in contrast to CMG2^{insCC} and CMG2^{delT}, ectopically expressed CMG2^{insC} is normally trafficked through the biosynthetic pathway, and reaches the cell surface where it can bind a surrogate ligand, altogether indicating that the protein is at least partially functional. We therefore investigated whether blocking the NMD pathway in patient fibroblasts, to inhibit mRNA degradation, would lead to a partial rescue of CMG2^{insC}. We chose to silence the expression of Upf1, a member of the group I RNA helicase and ATPase family (Bhattacharya et al, 2000), which is crucial for the initiation of the NMD pathway (Isken et al, 2008). Knock down of *upf1* was recently shown to rescue the phenotypes of Ulrich Disease and long-QT Syndrome in fibroblasts (Gong et al, 2011; Usuki et al, 2006). Upf1 expression could be efficiently silenced by siRNA in fibroblasts from patients 1, 4 and 5, which harbor the insC, insCC and delT respectively, as well as in fibroblasts from a healthy control (Fig. 6A). Remarkably, CMG2^{insC} could be detected by Western blotting (Fig. 6B) and by FACS analysis (Fig. 6CD) upon Upf1 silencing in Patient 1 cells. Due to the homozygosity of Patient 1 and the low level of *cmg2* mRNA expressed in the fibroblasts, we could detect a significant increase in *cmg2* mRNA upon Upf1 silencing (Fig. S2). Consistent with the ectopic expression studies (Fig. 2), CMG2^{insC} rescued by Upf1 silencing was properly targeted to the cell surface of fibroblasts (Fig. 6CD). No rescue of CMG2^{insCC} and CMG2^{delT} could be observed upon Upf1 silencing in cells from patients 4 and 5 (Fig. 6B), as expected from their efficient targeting to ERAD.

Conclusion

The present analysis of the HFS frameshift mutations in exon 13 allows us to draw several interesting conclusions. Firstly, it highlights the importance of depth analyses of the molecular consequences of the specific mutations found in a given patient, be it HFS or another monogenic disease, in order to evaluate the potential therapeutic targets. Indeed not only the site of a mutation is crucial but the exact nature of the insertion/deletion, i.e. the change in reading frame. Secondly, our study underlines that the genotype-phenotype analyses should be performed at the mRNA, the protein and the functional level, before significant conclusions can be drawn in term of therapeutic strategy and targets. And finally, more specific to HFS, our study shows that the HFS c.1073-1074insCC and c.1074delT mutations lead to defects not only at the mRNA level but also at the protein folding level. This leave little hope for a chemical based therapy to rescue CMG2 expression levels and thereby function recovery, leaving gene replacement as the only possibility in these patients. In contrast, a potential drug based therapy is not to be excluded for patient carrying the c.1073-1074insC mutation, fortunately the most frequent of the three frameshift mutations (Deuquet et al, 2011b). Indeed our study indicates that rescuing the *cmg2* mRNA in these patients allows the synthesis of a viable protein and point to the NMD pathway as a potential therapeutic target. Consistent with the differences we observe in terms of protein properties between CMG2^{insC} and CMG2^{insCC/delT}, a survey of the literature indicates that patients harboring the c.1073-1074insCC or c.1074delT mutations always suffer from the severe form of the disease, while more moderate symptoms are reported for patients with the c.1073-1074insC mutations, further raising the hope that even limited mRNA rescue could have an impact on the severity of the symptoms.

MATERIALS AND METHODS

Cell culture, plasmids, transfections and real time q-PCR

Primary patient-derived fibroblasts were grown in Dulbecco's Modified Eagle Medium (DMEM) medium (Gibco) at 37 °C complemented with fetal bovine serum (FBS) and penicillin and streptomycin. HeLa cells (ATCC) were grown in complete Modified eagles medium (MEM) (Sigma) at 37 °C supplemented with fetal bovine serum (FBS), L-glutamine, non-essential amino acids (NEAA), and penicillin and streptomycin. The human CMG2 (isoform 4, Swiss-Port P58335-4, GenBank AK091721.1) gene, tagged with a V5 epitope at the C-terminus, was cloned in the pcDNA3.1/V5-HIS- TOPO expression vector (Invitrogen). Mutations were generated with the QuickChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA). The full-length cDNAs encoding EGFP and human CD4 was cloned from pEGFP-C2 (Invitrogen) and pCD4-CMX (a gift from D. Trono), respectively, into the pcDNA3.1/V5-HIS-TOPO expression vector (Invitrogen). To construct the plasmids expressing chimeric protein EGFP-tail^{mut} or CD4-tail^{mut}, the full-length EGFP, the 1254bp from 5' end of CD4 and the 156bp from 3' end of CMG2 delT were cloned. The EGFP or CD4 fragment was then ligated with the fragment corresponding to the mutant tail with an NdeI cutting site designed at one end of them. The final cDNAs expressing the chimeric protein was inserted into the pcDNA3.1/V5-HIS-TOPO expression vector (Invitrogen). The 49 amino acids corresponding to the CMG2delT cytosolic tail were synthesized by EZBiolab. For the real-time PCR, RNA was extracted from a 5 cm dish using the RNeasy kit (Qiagen). 1 mg of the total RNA extracted was used for the reverse

transcription using random hexamers and superscript II (Invitrogen). A 1:40 dilution of the cDNA was used to perform the real-time PCR using SyBr green reagent (Roche). mRNA levels were normalized using three housekeeping genes: TATA-binding protein (TBP), β -microglobulin and β -glucuronidase (GUS).

Toxin, Antibodies and Reagents

The toxin subunit PA and biotin-labeled PA were a gift from S. Leppla. The anti-PA antibody was purchased from List Biological Laboratories, Inc. The anti-human CMG2 monoclonal antibodies 2F6, generated in our lab, was previously described (Deuquet et al, 2011a). Polyclonal antibody against calnexin was from Eurogentec, polyclonal goat anti human CMG2 and polyclonal goat anti human CD4 antibodies were from R&D Systems, anti-ubiquitin antibody from Santa Cruz, anti-EGFP-HRP antibody from the D. Trono laboratory. Protein G-agarose-conjugated beads and Protein A-agarose-conjugated beads were from GE Healthcare. MG132, from Sigma and Bortezomib, from Santa Cruz, were dissolved in DMSO and used at a final concentration of 10 μ M in complete medium. Bafilomycin was from Sigma and was dissolved in absolute ethanol and used at a final concentration of 0.1 μ M. Leupeptin was purchased from Roche, dissolved in water and used at a final concentration of 250 μ g/ml. Bis-ANS (4,4'-dianilino-1,1'-binaphthyl-5,5'-disulfonic acid, dipotassium salt) was from Invitrogen, dissolved in water and used at a final concentration of 50 μ M. Treatments with EndoH were performed as previously described (Abrami et al, 2006). The fixation reagent BD CellFix and permeabilization reagent FACS Permeabilizing Solution 2 were from BD BioScience.

Immunoprecipitation and RNAi experiments

siRNA of human *upf1* was purchased from Qiagen. As control siRNA we used the following target sequence of the viral glycoprotein VSV-G: 5'attgaacaaacgaacaagga 3'. For gene silencing, transfections of 100nM siRNAs were carried out with Lipofectamine 2000 (Invitrogen) and the cells were analyzed 72h after transfection. For immunoprecipitation, cells were washed three times with PBS and then were lysed by incubation for 30 min at 4 °C with 0.5 % NP-40, 500 mM Tris-HCl (pH 7.4), 20 mM EDTA, 10 mM NaF, 30 mM sodium pyrophosphate decahydrate, 2 mM benzamidine, 1 mM PMSF, 1 mM NEM and a cocktail of protease inhibitors (Roche). Cell lysates were centrifuged for 3 min at 2000 g and the supernatants were precleared with protein G-agarose conjugated beads and were incubated for 16 h at 4°C with antibodies and beads. The beads were washed three times with the same lysis buffer and were resuspended in sample buffer (4 x). The samples were heated at 95°C for 5 min and migrated on SDS-PAGE. Western blotting was performed with the iBlot (Invitrogen) according to the manufacturers instructions. Quantification of the blots was done using the Typhoon Imager (Image QuantTool, GE healthcare).

Metabolic labeling

For metabolic labeling, Hela cells were transfected for 24 h with respective cDNA and cells were starved with methionine/cysteine-free medium, incubated for a 5 min or 25 min pulse at 37 °C with 50 μ Ci/ml [³⁵S] methionine (Hartman Analytics). For chasing experiments, already pulsed cells were incubated for different chase time period at 37 °C in complete medium with a 10-fold excess of non-radioactive methionine and cysteine. Cells were then lysed with lysis buffer described above and immunoprecipitations were performed with cell lysates using protein A-agarose conjugated beads and respective antibodies. Samples were then migrated on SDS-PAGE after which the gel was incubated with a fixative solution (25 % isopropanol, 65 % H₂O, 10 % acetic acid), followed by a 30 min incubation with signal enhancer Amplify NAMP100 (Amersham). The dried gels were exposed to a Hyperfilm MP (Amersham) or quantified with the Typhoon Imager (Image QuantTool, GE healthcare).

Molecular dynamics simulations

Molecular dynamics (MD) simulations were used to further characterize the CMG2 cytoplasmic domain. As a first step, delT sequence and the corresponding section of the wild type (i.e. P342—S401) were modelled as a random coil. To explore efficiently the conformational space of these peptides, the MD simulations were carried out in the generalized Born implicit solvent. This solvent approximation, where the water molecules are described as a dielectric continuum, considerably reduces the computational time, allowing to reach hundreds of ns long simulations. In a second step, the equilibrated cytoplasmic tails were linked to CMG2 helical TM domain. The resulting proteins were inserted and equilibrated in a 80x80 Å² Palmitoyl Oleoyl Phosphatidyl Choline (POPC) membrane patch (Humphrey et al, 1996) to characterize its structure in a phospholipid bilayer.

All simulations were performed using NAMD (Phillips et al, 2005) engine, with the CHARMM27 force field (Brooks et al, 2009), including CMAP corrections. For the explicit solvent simulations, TIP3P water (Jorgensen et al, 1983) parameterization was used to describe the water molecules. The spatial overlapping of lipid molecules and protein were removed and the resulting protein-membrane system was solvated in variable-size water box, neutralized through the addition of NaCl at a concentration of 150 mM. The

periodic electrostatic interactions were computed using the particle-mesh Ewald (PME) summation with a grid spacing smaller than 1 Å. All systems were first minimized by 2000 conjugate gradient steps. All systems were subsequently gradually heated from 0 to 300 K in 800 ps with a constraint on the protein backbone scaffold. Finally, the systems were equilibrated for 10 ns at 300K. Free molecular dynamics of all equilibrated system were performed with a 2 fs integration time step using the RATTLE algorithm applied to all bonds. Constant temperature (300K) was imposed by using Langevin dynamics (Brunger & Brooks, 1984), with damping coefficient of 1.0 ps. In explicit solvent simulations, constant pressure of 1 atm was maintained with a Langevin piston dynamics (Feller et al, 1995), 200 fs decay period and 50 fs time constant. First, we performed 300 ns-long simulations for the characterization of each of the cytoplasmic tails folding in implicit solvent. Two simulations, lasting 100 ns each, were then carried out on the model of CMG2 transmembrane and cytoplasmic tail.

Hydrophobicity plot, disorder tendency and secondary structure prediction

The hydrophobicity of sequences corresponding to the cytosolic tail of WT CMG2 or delT mutant (starting from A359 and H359 for WT and delT respectively) was analyzed using the online ExPASy tool ProtScale (<http://web.expasy.org/protscale/>). The disorder tendency was predicted with IUPred (<http://iupred.enzim.hu/>). All data were extracted and plotted with Prism graphpad 5. No structural data are currently available for CMG2 wild type and delT cytoplasmic domain. We used secondary structure prediction tools to estimate the structural content of these domains: HNN, Jpred, NetSurfP, PSIPred, ProteinPredict. The final estimated structure was obtained by averaging the output of these different algorithms.

Circular Dichroism (CD)

CD experiments were carried out at room temperature on a Jasco J-815 spectrometer. Cells with 0.1cm path length were used. The synthetic peptide of CMG2 delT was diluted to a concentration of 20 µM in a buffer containing 20 mM Na₂HPO₄ (pH 7.4) and 40 % of acetonitrile. The buffer spectrum was subtracted for each sample.

Flow Cytometric Analysis

To permeabilize the cells, cells were fixed with BD CellFix reagent for 15 min and permeabilized by FACS Permeabilizing Solution 2 for 20 min at room temperature. Either permeabilized or non-permeabilized cells were incubated with anti-hCMG2 antibody for 30 min on ice. And after two times of wash with PBS + 20 mM EDTA + 2 % FCS, the cells were then incubated with a fluorescent secondary antibody for another 30 min on ice. Cells were washed for three times before they were fixed in the end. The stained cells were evaluated with Accuri C6 (BD BioScience). FACS data were analyzed using FlowJo software (FlowJo).

ANS binding assay

Binding of ANS was measured in 96-well plate with the spectrofluorimeter SpectraMax plate reader (Molecular Device). An excitation wavelength of 360 nm and an emission wavelength of 470 nm were applied. Synthetic peptide or purified protein was dissolved or diluted to a final concentration of 0.1 mg/ml with 20 mM sodium phosphate, pH 7.4 and 2 % acetonitrile. The final concentration of ANS was 50 µM.

Immunofluorescence Microscopy

Hela cells were fixed with 3% paraformaldehyde for 20 min and labelled with an anti-hCMG2 antibody followed by Alexa-conjugated anti-rat (555 nm) and Alexa- conjugated streptavidin antibody (647 nm). The nuclei were stained by Hoeschst dye. Images were acquired using a 63x lens on LSM-710 Laser scanning microscope (Carl Zeiss Microimaging, Inc.). The Fiji software was used for the processing of the images.

ACKNOWLEDGEMENTS

We are very grateful to the patients and their parents for accepting to be part of this study and for their collaborative attitude. We thank the Associazione I.S.I (www.associazioneisi.it) for linking patients, and their families, suffering for HFS. We thank Béatrice Kunz for performing the qPCR analysis, Ioan Iacovache and Bruno Fauvet for the circular dichroism, Miguel Garcia and his platform for helping with the FACS analysis, Asvin Lakkaraju and Sarah Friebe for the microscopy. This work was supported by the Swiss National Science Foundation (G.v.d.G.), the Fondation Telethon Action Suisse, the Fondation S.A.N.T.E.-Vaduz/Aide au soutien des nouvelles thérapies, the Gerbert Rûf Stiftung, Associazione I.S.I and the Fondation Minkoff.

AUTHOR CONTRIBUTIONS

Shixu Yan designed and performed the majority of the experiments and wrote the paper, Thomas

Lemmin, and his supervisor Matteo Dal Perarro, did the molecular modeling, Suzanne Salvi did all the maintenance of patient cell cultures, Ekkehart Lausch and Dariusz Rokicki are the physician of some the patients, Andrea Superti-Furga edited the paper and provided medical background on the disease, and F. Gisou van der Goot designed the study, the experiments and wrote the paper.

FOR MORE INFORMATION:

<http://www.orpha.net/>

<http://www.ncbi.nlm.nih.gov/omim/236490>

<http://www.ncbi.nlm.nih.gov/omim/228600>

<http://www.associazioneisi.it/>

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

- Abrami L, Bischofberger M, Kunz B, Groux R, van der Goot FG (2010) Endocytosis of the anthrax toxin is mediated by clathrin, actin and unconventional adaptors. *PLoS pathogens* **6**: e1000792
- Abrami L, Leppla SH, van der Goot FG (2006) Receptor palmitoylation and ubiquitination regulate anthrax toxin endocytosis. *The Journal of cell biology* **172**: 309-320
- Aoyagi T, Miyata S, Nanbo M, Kojima F, Matsuzaki M (1969) Biological activities of leupeptins. *The Journal of antibiotics* **22**: 558-568
- Bell SE, Mavila A, Salazar R, Bayless KJ, Kanagala S, Maxwell SA, Davis GE (2001) Differential gene expression during capillary morphogenesis in 3D collagen matrices: regulated expression of genes involved in basement membrane matrix assembly, cell cycle progression, cellular differentiation and G-protein signaling. *Journal of cell science* **114**: 2755-2773
- Bhattacharya A, Czaplinski K, Trifillis P, He F, Jacobson A, Peltz SW (2000) Characterization of the biochemical properties of the human Upf1 gene product that is involved in nonsense-mediated mRNA decay. *RNA* **6**: 1226-1235
- Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoseck M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* **30**: 1545-1614
- Brunger A, Brooks CL (1984) Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chemical Physics Letters* **105**: 495-500
- Deuquet J, Abrami L, Difeo A, Ramirez MC, Martignetti JA, van der Goot FG (2009) Systemic hyalinosis mutations in the CMG2 ectodomain leading to loss of function through retention in the endoplasmic reticulum. *Human mutation* **30**: 583-589
- Deuquet J, Lausch E, Guex N, Abrami L, Salvi S, Lakkaraju A, Ramirez MC, Martignetti JA, Rokicki D, Bonafe L, Superti-Furga A, van der Goot FG (2011a)
- Hyaline fibromatosis syndrome inducing mutations in the ectodomain of anthrax toxin receptor 2 can be rescued by proteasome inhibitors. *EMBO molecular medicine* **3**: 208-221
- Deuquet J, Lausch E, Superti-Furga A, van der Goot FG (2011b) The dark sides of capillary morphogenesis gene 2. *The EMBO journal* **31**: 3-13
- Dosztanyi Z, Csizmok V, Tompa P, Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**: 3433-3434
- Dowling O, Difeo A, Ramirez MC, Tükel T, Narla G, Bonafe L, Kayserili H, Yuksel- Apak M, Paller AS, Norton K, Teebi AS, Grum-Tokars V, Martin GS, Davis GE, Glucksman MJ, Martignetti JA (2003) Mutations in capillary morphogenesis gene-2 result in the allelic disorders juvenile hyaline fibromatosis and infantile systemic hyalinosis. *American journal of human genetics* **73**: 957-966
- Drose S, Altendorf K (1997) Bafilomycins and concanamycins as inhibitors of V- ATPases and P-ATPases. *The Journal of experimental biology* **200**: 1-8
- Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nature Reviews: Molecular Cell Biology* **6**: 197-208
- Engel J, Chiquet M (2011) An overview of extracellular matrix structure and function. In *The extracellular Matrix: an overview*, Mecham RP (ed), 1, pp 1-39.
- Feller SE, Zhang Y, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation: the Langevin piston method. *The Journal of Chemical Physics* **103**: 4613
- Gasteiger E, HC, Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. (2005) Protein Identification

and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, Humana Press, Walker JM (ed), pp 571-607

Gong Q, Stump MR, Zhou Z (2011) Inhibition of nonsense-mediated mRNA decay by antisense morpholino oligonucleotides restores functional expression of hERG nonsense and frameshift mutations in long-QT syndrome. *Journal of molecular and cellular cardiology* **50**: 223-229

Hanks S, Adams S, Douglas J, Arbour L, Atherton DJ, Balci S, Bode H, Campbell ME, Feingold M, Keser G, Kleijer W, Mancini G, McGrath JA, Muntoni F, Nanda A, Teare MD, Warman M, Pope FM, Superti-Furga A, Futreal PA, Rahman N (2003) Mutations in the gene encoding capillary morphogenesis protein 2 cause juvenile hyaline fibromatosis and infantile systemic hyalinosis. *American journal of human genetics* **73**: 791-800

Hegde RS, Ploegh HL (2010) Quality and quantity control at the endoplasmic reticulum. *Current opinion in cell biology* **22**: 437-446

Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* **14**: 33-38, 27-38

Iacovache I, Degiacomi MT, Pernot L, Ho S, Schiltz M, Dal Peraro M, van der Goot FG (2011) Dual chaperone role of the C-terminal propeptide in folding and oligomerization of the pore-forming toxin aerolysin. *PLoS pathogens* **7**: e1002135

Isken O, Kim YK, Hosoda N, Mayeur GL, Hershey JW, Maquat LE (2008) Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. *Cell* **133**: 314-327

Jackson MR, Nilsson T, Peterson PA (1990) Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum. *The EMBO journal* **9**: 3153-3162

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**: 926

Kisselev AF, Goldberg AL (2001) Proteasome inhibitors: from research tools to drug candidates. *Chemistry & biology* **8**: 739-758

Liu S, Leppla SH (2003) Cell surface tumor endothelium marker 8 cytoplasmic tail- independent anthrax toxin binding, proteolytic processing, oligomer formation and internalization. *J Biol Chem* **278**: 5227-5234

Nilsson T, Jackson M, Peterson PA (1989) Short cytoplasmic sequences serve as retention signals for transmembrane proteins in the endoplasmic reticulum. *Cell* **58**: 707-718

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**: 1781-1802

Rebbapragada I, Lykke-Andersen J (2009) Execution of nonsense-mediated mRNA decay: what defines a substrate? *Current opinion in cell biology* **21**: 394-402

Shieh JT, Swidler P, Martignetti JA, Ramirez MC, Balboni I, Kaplan J, Kennedy J, Abdul-Rahman O, Enns GM, Sandborg C, Slavotinek A, Hoyme HE (2006) Systemic hyalinosis: a distinctive early childhood-onset disorder characterized by mutations in the anthrax toxin receptor 2 gene (ANTRX2). *Pediatrics* **118**: e1485-1492

Shin J, Dunbrack RL, Jr., Lee S, Strominger JL (1991) Signals for retention of transmembrane proteins in the endoplasmic reticulum studied with CD4 truncation mutants. *Proceedings of the National Academy of Sciences of the United States of America* **88**: 1918-1922

Slavik J (1982) Anilino-naphthalene sulfonate as a probe of membrane composition and function. *Biochimica et biophysica acta* **694**: 1-25

Stucki U, Spycher MA, Eich G, Rossi A, Sacher P, Steinmann B, Superti-Furga A (2001) Infantile systemic hyalinosis in siblings: clinical report, biochemical and ultrastructural findings, and review of the literature. *American journal of medical genetics* **100**: 122-129

Tomba P (2012) Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci*

Urbina F, Sazunic I, Murray G (2004) Infantile systemic hyalinosis or juvenile hyaline fibromatosis? *Pediatric dermatology* **21**: 154-159

Usuki F, Yamashita A, Kashima I, Higuchi I, Osame M, Ohno S (2006) Specific inhibition of nonsense-mediated mRNA decay components, SMG-1 or Upf1, rescues the phenotype of Ullrich disease fibroblasts. *Molecular therapy : the journal of the American Society of Gene Therapy* **14**: 351-360

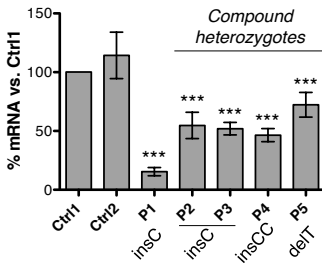
van der Goot G, Young JA (2009) Receptors of anthrax toxin and cell entry. *Molecular aspects of medicine* **30**: 406-412

FIGURE LEGENDS

A



B



C

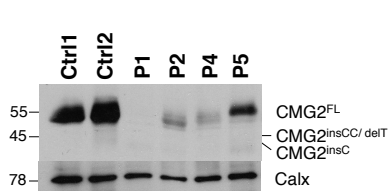
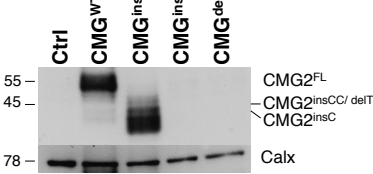
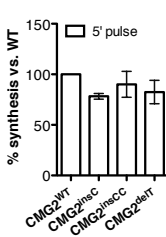


Figure 1: Analysis of frameshift mutations in HFS patients. A: amino acid sequence alignment of CMG2^{WT}, CMG2^{insCC}, CMG2^{delT} and CMG2^{insC} generated with Jalview. Sequences were highlighted according to the percentage of identity. B: mRNA levels of *cmg2* in patient-derived fibroblasts were analyzed by quantitative RT-PCR. The values representing each mRNA level were plotted as the percentage of control 1 (ctrl1) fibroblast. Error bars represent standard errors (n>3). Two-tailed paired t-test analysis between ctrl1 and patient-derived fibroblasts was performed. *** : p < 0.001. C: 20 μ g of total cell extracts from patient-derived fibroblasts were analyzed by SDS- PAGE under non-reducing condition and Western blotting using the 2F6 anti-hCMG2 antibody.

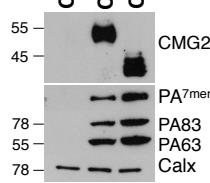
A



B



D



C

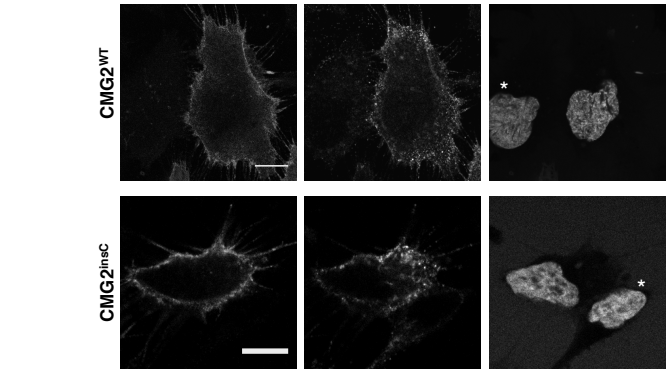


Figure 2: Ectopic expression of CMG2^{insC}, CMG2^{insCC} and CMG2^{delIT} A: HeLa cells were transfected or not (ctrl) for 48 h with cDNA expressing either WT or mutant CMG2. 20 μ g of total cell extracts were analyzed by SDS-PAGE under non-reducing condition and Western blotting using the 2F6 anti-hCMG2 and anti-calnexin antibodies. B: HeLa cells were transfected or not (ctrl) for 24 h with cDNA expressing either WT or mutant CMG2. Cells were pulsed for 5 min with [³⁵S] cysteine/methionine. Immunoprecipitation was performed using anti-hCMG2 antibody and samples were analyzed by SDS-PAGE and autoradiography. Quantification was performed using the Typhoon Imager. Error bars represent standard errors (n=3). Two-tailed paired t-test analysis was performed. *** : p < 0.001. C: *tem8* knocked-down HeLa cells were transfected or not (ctrl) for 48 h with cDNA expressing either WT or *insC* CMG2. Cells were incubated with anthrax PA at a final concentration of 500 ng/ml on ice for 1h and then shifted to 37° C for 20 min before they are lysed. 20 μ g of total cell extracts were analyzed by SDS-PAGE under non-reducing or reducing condition and Western blotting using the 2F6 anti-hCMG2 antibody, anti-PA and anti-calnexin antibodies. The dotted lines represent the non-transfected control cells and the solid lines CMG2^{WT} or CMG2^{insC}. D: *tem8* knocked-down HeLa cells grown on coverslips were transfected or not (ctrl) for 48 h with cDNA expressing either WT or CMG2^{insC}. Cells were incubated with biotin-labeled anthrax PA at a final concentration of 1 μ g/ml on ice for 1h and then shifted to 37°C for 5 min before fixation and labeling with anti-hCMG2 and streptavidin antibodies and Hoechst. The scale bars represent 10 μ m.

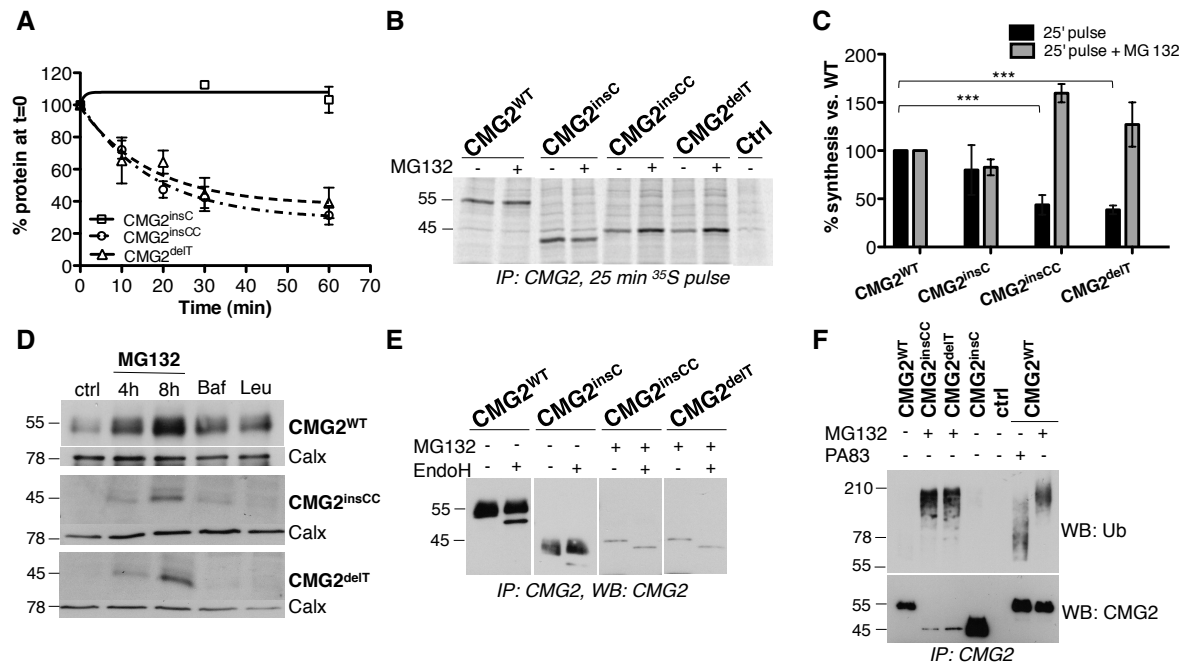


Figure 3: CMG2^{insCC} and CMG2^{delIT} are powerful ERAD substrates. A: HeLa cells were transfected or not (ctrl) for 24 h with cDNA expressing either WT or mutant CMG2. Cells were pulsed for 5 min with [³⁵S] cysteine/methionine and then chased for different time periods. Immunoprecipitation was performed using anti-hCMG2 antibody and samples were analyzed by SDS-PAGE and autoradiography. Quantification was performed using the Typhoon Imager. For each experiment, the intensities of signals were normalized to that of the same cDNA at time t = 0 min. Error bars represent standard errors (n=3). The curve was smoothed by nonlinear fit according to the one phase decay model using the software Prism graphpad 5. B: HeLa cells were transfected or not (ctrl) for 24 h with cDNA expressing either WT or mutant CMG2. Cells were treated with 10 μ M MG132 or not for 4 h prior to being pulsed for 25 min with [³⁵S] cysteine/methionine. Immunoprecipitation was performed using anti-hCMG2 antibody and samples were analyzed by SDS-PAGE and autoradiography. C: Quantification for WT or mutant CMG2 with or without MG132 treatment in the 25 min radiolabeling experiments was performed using the Typhoon Imager. Error bars represent standard errors (n=3). D: HeLa cells were transfected or not (ctrl) for 48 h with cDNA expressing either WT or mutant CMG2. Cells were treated or not (ctrl) with 10 μ M MG132 for 4 h or 8 h, or with 100 μ M Bafilomycin for 4 h, or with 250 μ g/ml leupeptin for 16 h. 20 μ g of total cell extracts were analyzed by SDS-PAGE under non-reducing condition and Western blotting using the 2F6 anti-hCMG2 antibody and anti-calnexin antibody. E: HeLa cells were transfected or not (ctrl) for 48 h with cDNA expressing either WT or mutant CMG2 and were treated or not with 10 μ M MG132 for 4 h. Cell lysates were subjected to immunoprecipitation with an anti-hCMG2 antibody. The products from

immunoprecipitation were treated with or without EndoH and were analyzed by SDS-PAGE and Western blotting using anti-hCMG2 antibody. F: HeLa cells were transfected or not (ctrl) for 48 h with cDNA expressing either WT or mutant CMG2. Cells were either treated with or without 10 μ M MG132 for 4 h or treated with 500 ng/ml anthrax PA at 4°C for 1 h followed by at 37°C for 10 min. Cell lysates were immunoprecipitated with an anti-hCMG2 antibody and were subsequently analyzed by SDS-PAGE and Western blotting using anti-ubiquitin antibody and anti-hCMG2 antibody.

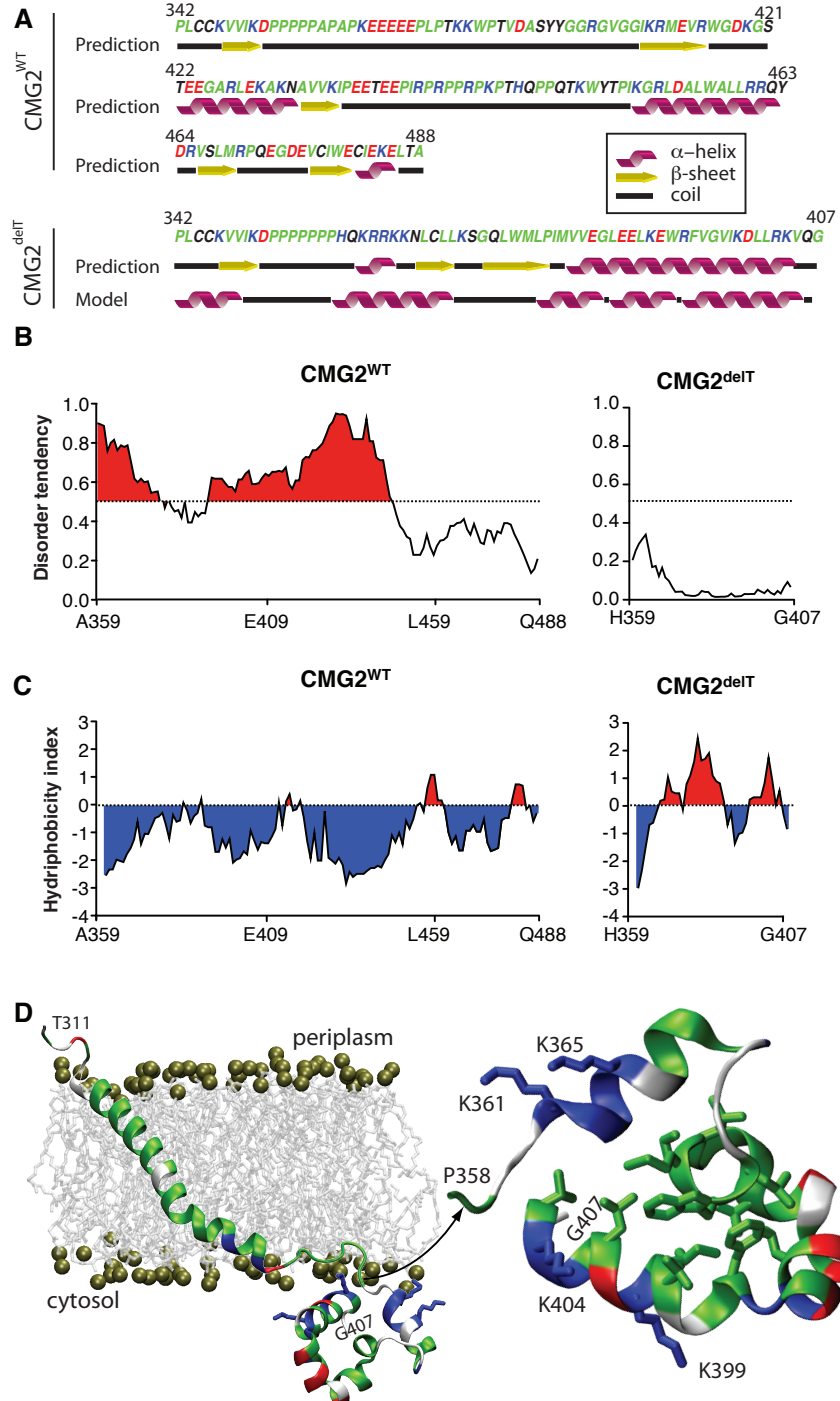


Figure 4: Structural predictions of the CMG2^{insCC/deIT} tail. A: Sequence and structural content for WT CMG2 and CMG2^{deIT} cytoplasmic tail. The sequences were color coded with polar residues in black, non polar in green, positively and negatively charged residues in red and blue respectively. The predicted secondary structure was the average output of five secondary structure prediction tools. B: Disorder tendency plot of the cytosolic tails of WT CMG2 and CMG2^{deIT} (starting from A359 and H359 for WT CMG2 and CMG2^{deIT} respectively). The curves' surfaces below the value of zero were colored in blue, representing the hydrophilic portion whereas the surfaces above zero in red, representing the

hydrophobic portion. C: Disorder prediction for the cytosolic tails of WT CMG2 and CMG2delT. The curves' surfaces above the threshold of 0.5 were colored in red, representing the predicted disordered portion. D: Molecular dynamics based model of CMG2delT cytosolic tail. The modeled structure was linked to CMG2 transmembrane (TM) domain. The protein was color coded as follows: polar residues in white, non polar in green, acidic and basic residues in red and blue respectively. The cytoplasmic hydrophobic core interacts with the membrane surface. Side chains of critical residues are shown in ball and sticks.

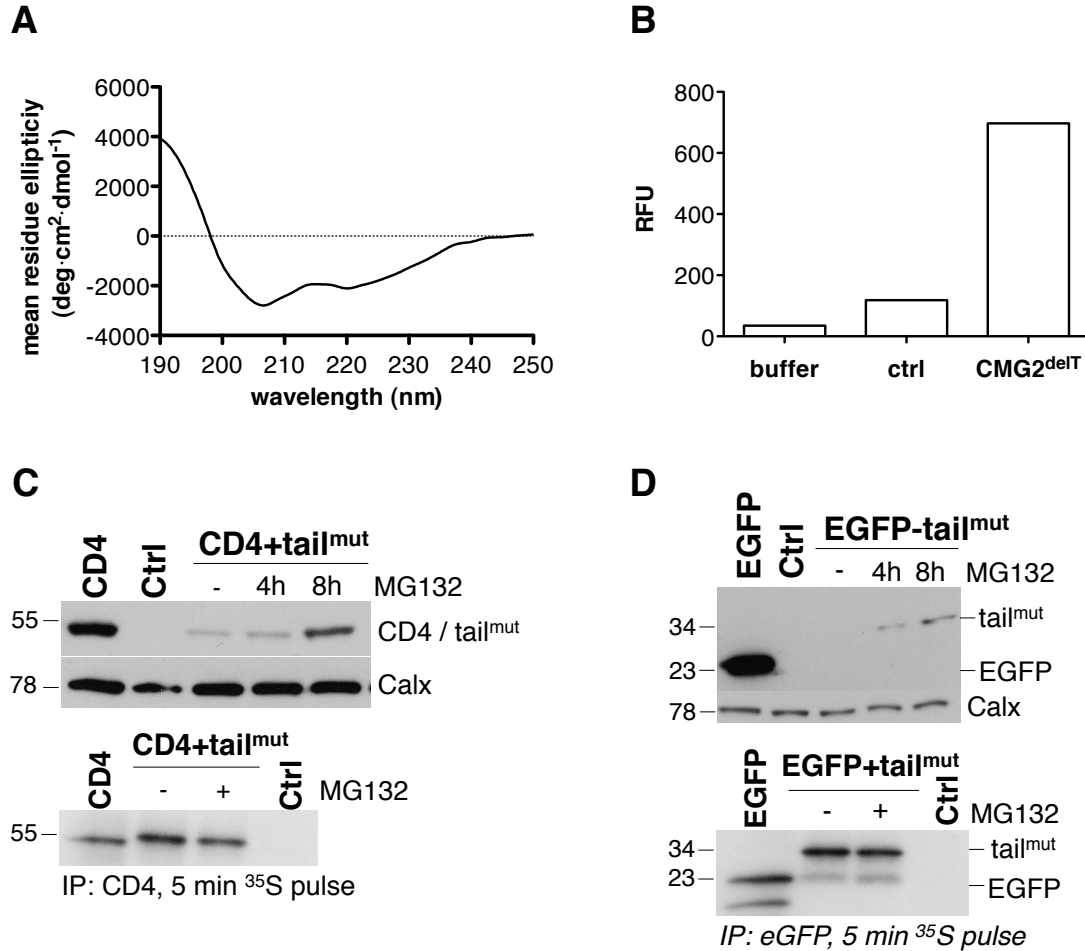


Figure 5: The CMG2insCC/deIT tail sequence is sufficient for targeting to degradation.

A: Far UV CD spectra of the cytoplasmic tails of CMG2delT. A synthetic peptide corresponding to the cytoplasmic tail of CMG2delT (starting from H359) was dissolved in 20 mM sodium phosphate (pH 7.4) with 40% acetonitrile and its far UV CD spectrum was analyzed with a Jasco J-815 spectrometer. B: Binding of the hydrophobic dye ANS to control (ctrl) and CMG2delT cytoplasmic tail. Synthesized peptides of ctrl and delT cytoplasmic tail were dissolved in 20 mM sodium phosphate (pH 7.4) with 2% acetonitrile to a final concentration of 0.1 mg/ml respectively. 50 μ M ANS was added to each sample and the relative fluorescence unit (RLU) was measured with a SpectraMax plate reader (Molecular Device). The graph showed was one representative result. C: HeLa cells were transfected or not (ctrl) for 48 h with cDNA expressing either WT EGFP or mutated tail (cytoplasmic tail of CMG2delT starting from H359) fused chimeric EGFP. Cells were treated or not with 10 μ M MG132 for 4 h or 8 h. 20 μ g of total cell extracts were analyzed by SDS-PAGE and Western blotting using anti-EGFP antibody and anti-calnexin antibody. To ensure that proteins were well synthesized, 24 h-transfected or not HeLa cells with or without 1 h 10 μ M MG132 treatment were pulsed for 5 min with [³⁵S] cysteine/methionine in parallel. Immunoprecipitation was performed using anti-EGFP antibody and the samples were analyzed by SDS-PAGE and autoradiography. D: HeLa cells were transfected or not (ctrl) for 48 h with cDNA expressing either WT CD4 or chimeric protein composed of the ectodomain and transmembrane part of CD4 and the mutated cytoplasmic tail of CMG2delT (starting from H359). Cells were treated or not with 10 μ M MG132 for 4 h or 8 h. 20 μ g of total cell extracts were analyzed by SDS-PAGE and Western blotting using anti-CD4 antibody and anti-calnexin antibody. To ensure that proteins were well synthesized, 24 h-transfected or not HeLa cells with or without 1 h 10 μ M MG132 treatment were pulsed for 5 min with [³⁵S] cysteine/methionine in parallel. Immunoprecipitation was performed using anti-CD4 antibody and the samples were analyzed by SDS-PAGE

and autoradiography.

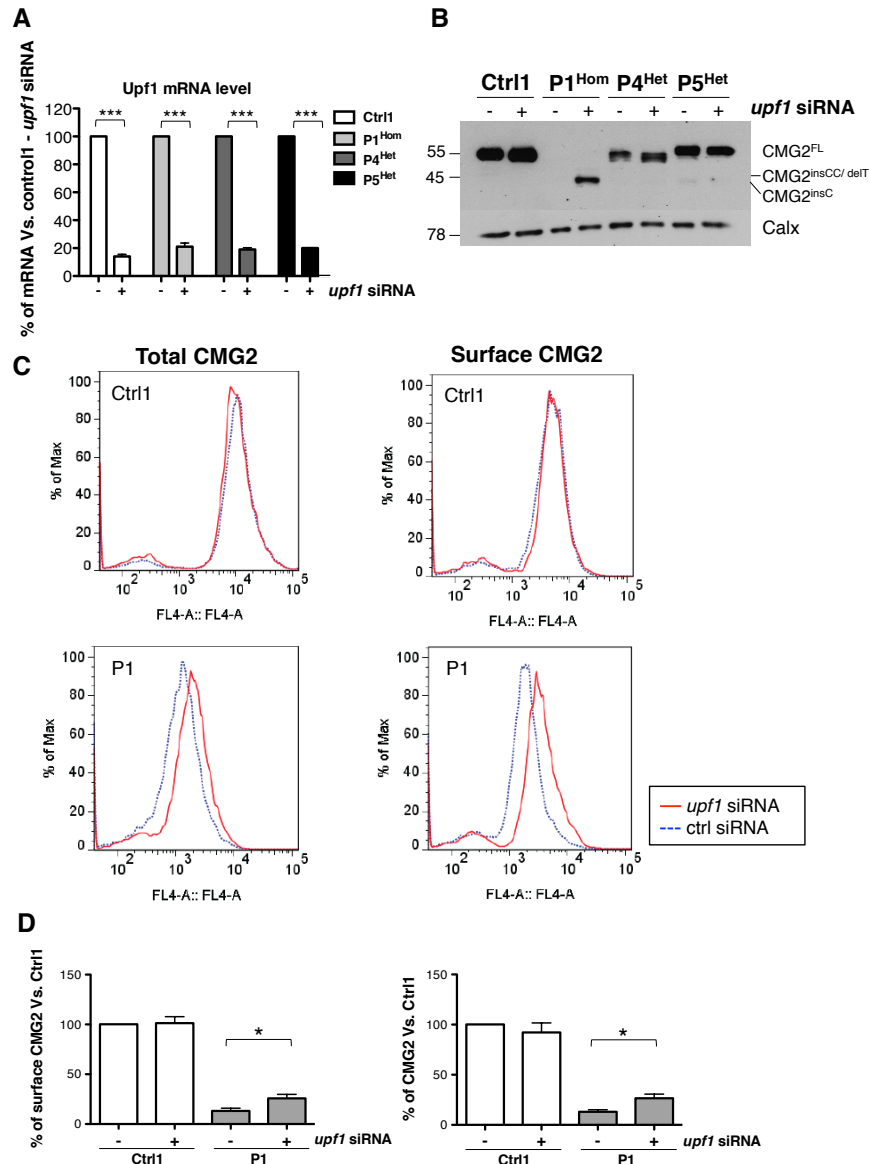


Figure 6: Inhibiting NMD pathway can rescue CMG2^{insC} in patient-derived fibroblasts.

A: mRNA levels of *upf1* in patient-derived fibroblasts after 72 h knocking down with or without siRNA targeting *upf1* were analyzed by quantitative RT-PCR. Under the condition that cells were without siRNA targeting *upf1*, a control siRNA against *vsvg* was used. The values representing each mRNA level were plotted as the percentage of fibroblast without siRNA targeting *upf1*. Error bars represent standard errors (n>4). Two-tailed paired t-test analysis between the same patient-derived fibroblasts with or without siRNA targeting *upf1* was performed. *** : p < 0.001. B: Ctrl1 and patient-derived fibroblasts were transfected with or without siRNA targeting *upf1* for 72 h. Under the condition that cells were without siRNA targeting *upf1*, a control siRNA against *vsvg* was used. 10µg of total cell extracts were analyzed by SDS-PAGE under non-reducing condition and Western blotting using anti-hCMG2 antibody and anti-calnexin antibody. C: Ctrl1 and Patient1 (P1) derived fibroblasts were transfected with or without siRNA targeting *upf1* for 72 h. Under the condition that cells were without siRNA targeting *upf1*, a control siRNA against *vsvg* was used. Both permeabilized and non-permeabilized cells were stained with anti-hCMG2 antibody for 30 min on ice followed by a secondary fluorescent antibody staining for 30 min. The cells were then washed, fixed and analyzed on a FACS analyzer Accuri C6 (BD BioScience). D: The mean fluorescence intensity of each sample from the FACS data shown above was extracted with the software FlowJo 7.6. Mean fluorescence intensities representing either the total amount of CMG2 expressed or the surface CMG2 were plotted as the percentage of that of Ctrl1 without siRNA targeting *upf1* respectively. Error bars represent standard errors (n=4). Two-tailed paired t-test analysis was performed. * : p < 0.05.

Table 1. HFS mutations analysed in this paper

Patient	HFS	Zygosity	DNA	Protein	Exon	References
1	Infantile	Homozygote	c.1073-1074insC	p.A359CfsX13	13	Deuquet et al, 2011
2	Infantile	Compound Heterozygote	c.566T>C	p.I189T	7	Dowling et al, 2003
			c.1073-1074insC	p.A359CfsX13	13	
3	Juvenile	Compound Heterozygote	c.928G>T	p.V310F	11	Deuquet et al, 2011
			c.1073-1074insC	p.A359CfsX13	13	
4	Infantile	Compound Heterozygote	c.566T>C	p.I189T	7	Hanks et al, 2003
			c.1073-1074insCC	p.A359LfsX51	13	
5	Infantile	Compound Heterozygote	c.116G>T	p.C39F	1	Deuquet et al, 2011
			c.1074delT	p.A359HfsX50	13	

Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence, according to the nomenclature proposed by the Human Genome Variation Society (www.hgvs.org/mutnomen). The initiation codon is codon 1.

Palmitoylated calnexin is a key component of the ribosome–translocon complex

Asvin KK Lakkaraju^{1,4}, Laurence Abrami^{1,4},
Thomas Lemmin², Sanja Blaskovic¹,
Béatrice Kunz¹, Akio Kihara³, Matteo Dal
Peraro² and Françoise Gisou van der Goot^{1,*}

¹Global Health Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, ²Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland and ³Faculty of Pharmaceutical Sciences, Hokkaido University, Sapporo, Japan

A third of the human genome encodes *N*-glycosylated proteins. These are co-translationally translocated into the lumen/membrane of the endoplasmic reticulum (ER) where they fold and assemble before they are transported to their final destination. Here, we show that calnexin, a major ER chaperone involved in glycoprotein folding is palmitoylated and that this modification is mediated by the ER palmitoyltransferase DHHC6. This modification leads to the preferential localization of calnexin to the perinuclear rough ER, at the expense of ER tubules. Moreover, palmitoylation mediates the association of calnexin with the ribosome–translocon complex (RTC) leading to the formation of a supercomplex that recruits the actin cytoskeleton, leading to further stabilization of the assembly. When formation of the calnexin–RTC supercomplex was affected by DHHC6 silencing, mutation of calnexin palmitoylation sites or actin depolymerization, folding of glycoproteins was impaired. Our findings thus show that calnexin is a stable component of the RTC in a manner that is exquisitely dependent on its palmitoylation status. This association is essential for the chaperone to capture its client proteins as they emerge from the translocon, acquire their *N*-linked glycans and initiate folding.
The EMBO Journal (2012) **31**, 1823–1835. doi:10.1038/emboj.2012.15; Published online 7 February 2012

Subject Categories: proteins

Keywords: calnexin; DHHC6; endoplasmic reticulum folding; palmitoylation

Introduction

More than a third of the human genome encodes for membrane proteins or proteins secreted into the extracellular milieu. These proteins are generally synthesized with an *N*-terminal signal sequence that targets the translating ribosome to the endoplasmic reticulum (ER) where it docks onto

the translocon (Braakman and Bulleid, 2011). The translocon is a multiprotein transmembrane complex composed of a central sec61 α tetrameric pore to which associate a number of accessory proteins such as sec61 β , sec62 or the TRAP (translocon-associated protein) complex (Skach, 2007). Once the ribosome is docked onto the translocon, synthesis resumes and the nascent protein is co-translationally translocated through the sec61 channel to the lumen of the ER. As the newly synthesized protein emerges from the translocon, it is handled sequentially by a variety of enzymes, in particular leading to the addition of *N*-linked glycans and the formation of disulphide bonds, which together increase the solubility and stability of the protein, as well as chaperones that promote folding by preventing aggregation of intermediates.

N-glycosylation is carried out by the oligosaccharyltransferase (OST), a hetero-oligomeric complex, which associates with the ribosome–translocon complex (RTC), allowing the addition of the branched oligosaccharide co- or post-translationally (Chavan and Lennarz, 2006). To prevent aggregation or too rapid folding, which might hinder *N*-glycosylation (McGinness and Morrison, 1994), the OST (Li *et al*, 2008) and the general chaperone BiP/Grp78 (Hammond and Helenius, 1994) maintain the polypeptide in a folding competent state.

Glycoproteins are subsequently handled by the lectin-binding chaperones calnexin and calreticulin (Aebi *et al*, 2010). Calnexin is a type I membrane protein and calreticulin is its soluble homologue. Proteins that undergo *N*-glycosylation acquire *en bloc* a glucose₃-mannose₉-*N*-acetylglucosamine₂ oligosaccharide. Calnexin/calreticulin can only bind monoglucosylated glycans. Oligosaccharide trimming is therefore required. This occurs immediately after the oligosaccharide has been transferred to the protein, first by ER glucosidase I, which removes the first glucose, and then by glucosidase II, which removes the second. The third glucose is also trimmed by glucosidase II, but its removal requires separation and repositioning of the enzyme with respect to the substrate (Deprez *et al*, 2005). This repositioning provides a time window during which the monoglucosylated polypeptide can interact with calnexin/calreticulin.

Glucose trimming must be rapid since calnexin was shown to bind proteins not only post-translationally but also co-translationally (Chen *et al*, 1995). It has been reported that upon signalling, MAP kinase and casein kinase 2-dependent phosphorylation of calnexin on Ser-563 allows its association with ribosomes (Chevet *et al*, 1999). We here investigated whether other mechanisms control the ability of calnexin to capture its substrates. Calnexin recently came up in four large-scale profiling studies of *S*-acylated proteins (Martin and Cravatt, 2009; Yang *et al*, 2010; Yount *et al*, 2010; Merrick *et al*, 2011). *S*-acylation is the addition of a fatty acid, generally C16, palmitate, but possibly also C18, stearate (Kordyukova *et al*, 2010), to cytosolic cysteine residues via a thioester bond (Linder and Deschenes, 2007). This modification is mediated by transmembrane palmitoyltrans-

*Corresponding author. Global Health Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), Station 15, Lausanne 1015, Switzerland. Tel.: +41 21 693 1791; Fax: +41 21 693 9538; E-mail: gisou.vandergoot@epfl.ch

⁴These authors contributed equally to this work

ferases harbouring a conserved cytosolic DHHC domain, 23 of which are present in the human genome (Greaves and Chamberlain, 2011). In validation of these profiling studies and in agreement with a study published during the revision of this manuscript (Lynes *et al*, 2011), we show that calnexin can be palmitoylated on two juxtamembranous cysteines and that this modification is mediated by a single palmitoyltransferase, DHHC6, out of the 16 DHHC enzymes present in the mammalian ER. S-acylation controls the localization of calnexin within the ER network, favouring its targeting to the central ER. While being a single copy organelle, the ER is indeed compartmentalized into a variety of still poorly characterized domains that allow it to fulfil its multiple functions, which in addition to protein folding include quality control, synthesis of most cellular lipids and storage/control of intracellular calcium (Pendin *et al*, 2011). The most apparent, morphologically distinct, ER regions are the highly branched tubular network that extends to the cell periphery, the dense perinuclear ER, which is formed by sheet-like structures, and the nuclear membrane (Puhka *et al*, 2007; Pendin *et al*, 2011). The two latter regions composed the rough ER, where the RTC complexes reside.

We found that calnexin is a stable component of the RTC and this association relies on calnexin S-acylation. Formation of the calnexin–RTC supercomplex leads to the recruitment of the actin cytoskeleton, which further stabilizes the assembly. The here-identified acylation-dependent interaction of calnexin with the RTC complex allows it to efficiently capture nascent glycosylated polypeptide chains, as they emerge from the translocon, and promote their folding.

Results

Calnexin is palmitoylated in the ER on two juxtamembrane cysteines

Four profiling studies (Martin and Cravatt, 2009; Yang *et al*, 2010; Yount *et al*, 2010; Merrick *et al*, 2011) identified, with high confidence, calnexin as an S-acylated protein in mammalian cells. By immunoprecipitation of either the endogenous or tagged calnexin from ³H-palmitate-labelled cells, we validated these profiling studies: the protein indeed incorporated radiolabelled palmitate, which could be removed by hydroxylamine hydrochloride treatment, indicating the involvement of a thioester bond (Figure 1A). This does not exclude that calnexin could be modified with either shorter or longer acyl chains as found for certain viral proteins (Kordyukova *et al*, 2010). ³H-palmitate incorporation occurred on both juxtamembranous cysteines (Figure 1B), although somewhat more efficiently on Cys-502, which is conserved in most species as well as in the testis-specific calnexin homologue, calmeglin (Supplementary Figure S1).

In contrast to other lipid modifications, S-acylation is reversible (Greaves and Chamberlain, 2011). Proteins such as Ras indeed undergo dynamic (<30 min) cycles of palmitoylation–depalmitoylation, and these cycles are crucial for their function (Rocks *et al*, 2010). In contrast to Ras, we found that ³H-palmitate turnover on calnexin is slow (Figure 1C). As a consequence, palmitoylated calnexin must accumulate in cells and we indeed found, as described below, that at steady state the vast majority of the protein is lipid modified. Palmitoylation of calnexin mostly occurs after it has been fully synthesized, since inhibition of protein synthesis with

cycloheximide 1 h prior to ³H-palmitate labelling only mildly affected palmitoylation (18 ± 4% decrease; Figure 1D).

Calnexin palmitoylation is mediated by DHHC6

We next identified the palmitoyltransferase that modifies calnexin. PCR analysis revealed that all DHHC enzymes, with the exception of DHHC15, 19 and 22, are expressed in HeLa cells (Supplementary Figure S2A). RNAi duplexes were screened for their efficiency in silencing these enzymes (Supplementary Table S1; Supplementary Figure S2A). Silencing of DHHC enzymes had minor effects (<25% change) on ³H-palmitate incorporation into endogenous calnexin with the exception of DHHC5 and DHHC6 (Figure 2A and B). Importantly, three different siRNA duplexes against DHHC6 led to a similar decrease in calnexin palmitoylation (Figure 2A) as did an shRNA construct against the non-coding region of DHHC6 (Supplementary Figure S2B). Recombination of shRNA-treated cells with a DHHC6 expression plasmid restored calnexin palmitoylation (Supplementary Figure S2B), indicating that the observed effect is not due to an off target effect of the shRNA or RNAi duplexes.

³H-palmitate incorporation experiments were also performed upon ectopic expression of calnexin–HA. Only DHHC6 silencing led to a marked decrease in palmitoylation (Figure 2C; Supplementary Figure S2C). We currently have no clear explanation as to why DHHC5 silencing affects palmitoylation of endogenous but not ectopically expressed calnexin. DHHC6 localizes to the ER (Supplementary Figure S2D and (Gorleku *et al*, 2011), consistent with the calnexin localization, while DHHC5 is mostly found in the Golgi apparatus (Supplementary Figure S2D). This however does not exclude its presence in the ER. Since DHHC6 is itself palmitoylated (personal communication; Gorleku *et al*, 2011), our findings raise the possibility—which will be tested in future studies—that DHHC6 palmitoylates DHHC6 or somehow regulates its function.

We next performed the reverse experiments, we monitored the effect of DHHC overexpression. An increase in palmitoylation of endogenous as well as ectopically expressed calnexin was observed upon DHHC6 overexpression but for none of the other enzymes (Figure 2D; Supplementary Figure S2E and F). Analysis of the two single cysteine calnexin mutants showed that both cysteines are modified by the same DHHC6 enzyme (Supplementary Figure S2E and F). Altogether, these observations show that DHHC6 modifies calnexin on both sites. Consistent with the ubiquitous expression of calnexin, we found that the DHHC6 expression is equally broad (Supplementary Figure S3A).

Since many of the experiments described in the present work involve DHHC6 silencing, we tested whether DHHC6 siRNA triggers the unfolded protein response. A gene profiling analysis using Affimetrix chips revealed that while slight changes in expression were observed for GADD45 (–1.3-fold), BiP (1.32-fold), XBP-1 (1.4-fold), HERP (1.5-fold), these changes are minor in comparison to what is observed upon UPR activation. Usually UPR induction leads to 4–27-fold increase in the expression levels of these genes at the mRNA level (Lee *et al*, 2003). Also, we did not observe significant XBP-1 splicing.

As a side note, DHHC6 overexpression had no effect on the palmitoylation of four other proteins known to be S-acylated that transit through the ER: the Transferrin receptor (Alvarez

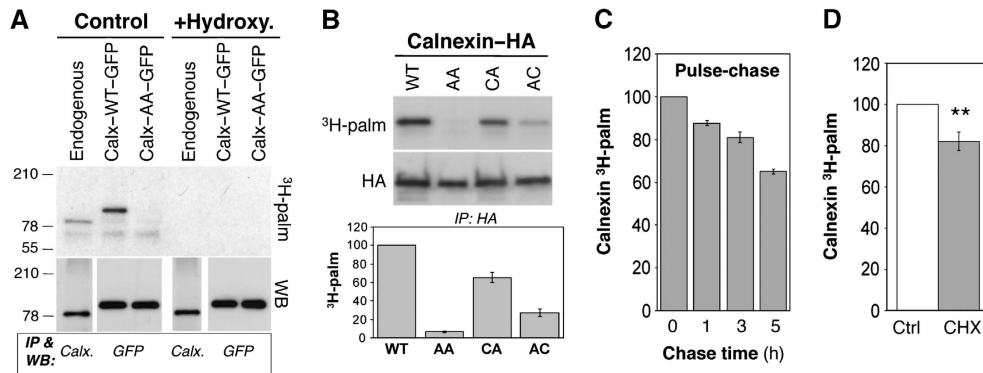


Figure 1 Calnexin is palmitoylated in the ER. (A) HeLa cells were transfected or not (endogenous) for 48 h with calnexin-WT-GFP or calnexin-AA-GFP. Cells were incubated with ³H-palmitic acid for 2 h. Cell extracts were treated or not with 1 M hydroxylamine hydrochloride pH 7.2 (1 h at room temperature) prior to immunoprecipitation using anti-calnexin or anti-GFP antibodies. Immunoprecipitates were split into two, run on SDS-PAGE and analysed by autoradiography or western blotting. (B) HeLa cells were transfected 24 h with calnexin-WT-HA or calnexin-CA-HA, calnexin-AC-HA or calnexin-AA-HA, labelled with ³H-palmitic acid and analysed as in (A) by autoradiography or western blotting (anti-HA). See Supplementary Figure S1 for positions of the cysteines. (Bottom panel) Autoradiograms were quantified using the Typhoon Imager. Errors correspond to standard deviations ($n = 5$). (C) Non-transfected HeLa cells were labelled as in (A). The cells were then washed and incubated in the normal medium and cell lysis was performed at different time points followed by immunoprecipitation as described in (A). Samples were analysed by autoradiography and the quantification was performed using the Typhoon Imager ($n = 3$). (D) HeLa cells were pretreated or not with cycloheximide (CHX) for 1 h followed by ³H-palmitic acid labelling in the presence or absence of the drug. Immunoprecipitation was performed using anti-calnexin antibodies. Samples were analysed by autoradiography and western blotting (anti-calnexin). Quantification was performed using the Typhoon Imager. Errors correspond to standard deviations ($n = 3$). ** $P < 0.01$.

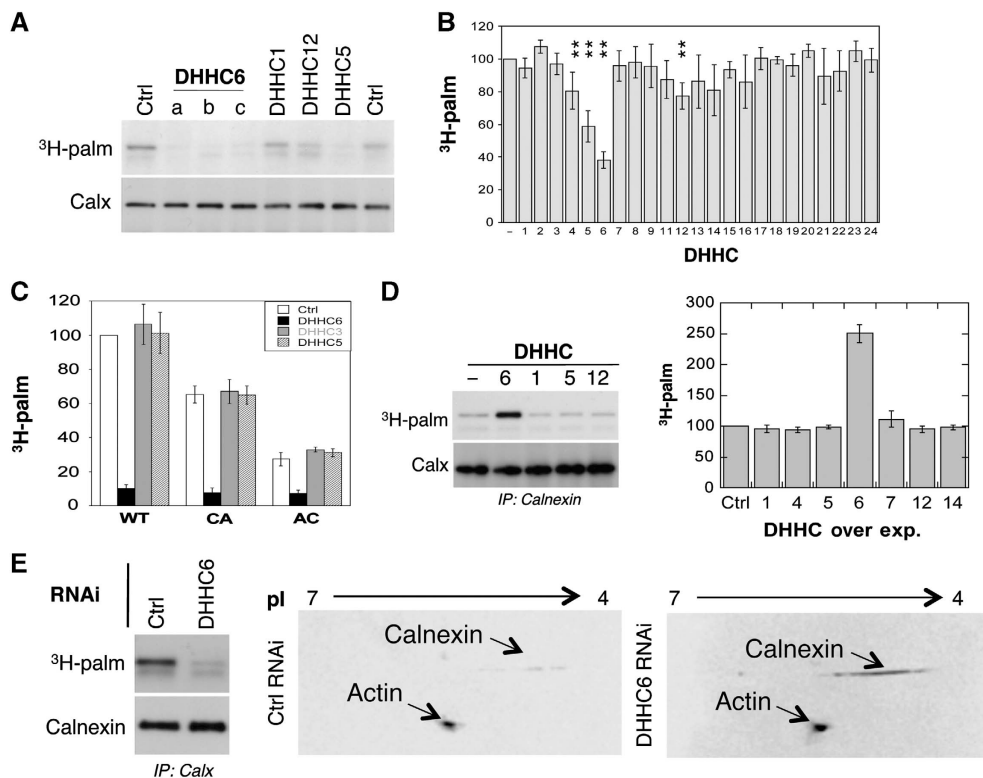


Figure 2 Calnexin is palmitoylated by DHHC6. (A) HeLa cells were transfected for 72 h with different DHHC siRNA or control siRNA (Ctrl). Cells were incubated with ³H-palmitic acid for 2 h prior to immunoprecipitation using anti-calnexin antibodies. Immunoprecipitates were split into two, run on SDS-PAGE and analysed by either autoradiography (³H-palmitate) or western blotting (anti-calnexin). (B) Autoradiograms and western blots from (A) were quantified using the Typhoon Imager. Errors correspond to standard deviations ($n = 4$). ** $P < 0.01$. For the identity of the siRNAs see Supplementary Table S1, for their efficiency see Supplementary Figure S2A. (C) HeLa cells were transfected 72 h with calnexin-WT-HA or calnexin-CA-HA, or calnexin-AC-HA and DHHC siRNAs, labelled with ³H-palmitic acid. Immunoprecipitates using anti-HA antibodies were split into two, run on SDS-PAGE gels and analysed by either autoradiography or western blotting (see Supplementary Figure S2C). Autoradiograms and western blotting were quantified using the Typhoon Imager. Errors correspond to standard deviations ($n = 4$). (D) HeLa cells were transfected 24 h with human DHHC cDNA, labelled with ³H-palmitic acid. Immunoprecipitates using anti-calnexin antibodies were analysed and quantified as in (B) ($n = 4$). For analysis of cysteine mutants see Supplementary Figure S2E and F. (E) HeLa cells were transfected for 72 h with DHHC6 siRNA or control siRNA (Ctrl). (Left panel) Cells were labelled with ³H-palmitic acid as analysed as in (A). (Right panel) In all, 50 μ g of cell extracts was analysed using 2D gel ZOOM IPGRunner system (Invitrogen) followed by western blotting using anti-calnexin and anti-actin antibodies.

et al, 1990), the Wnt co-receptor LRP6 (Abrami *et al*, 2008) and the two anthrax toxin receptors TEM8 and CMG2 (Abrami *et al*, 2006) (Supplementary Figure S3B), suggesting a certain specificity towards calnexin.

We next investigated what percentage of the calnexin population is S-acylated at steady state in HeLa cells. We previously found that glycosylphosphatidyl inositol-anchored proteins do not migrate in 2D gels unless their lipid moiety has been removed (Fivaz *et al*, 2000). We inferred that palmitoylation might also affect protein migration in 2D gels. Cell extracts from control and DHHC6 RNAi-treated HeLa cells were analysed by 2D-PAGE followed by western blotting against calnexin, and actin as an internal control. The analysis was performed 72 h after transfection with DHHC6 siRNA. This did not affect the total level of endogenous calnexin when compared with control cells, as evidenced by 1D SDS-PAGE (Figure 2E, left panel; also seen in Figure 2A, bottom blot). As a striking validation of our strategy, DHHC6 silencing led to a 9.1 ± 0.9 ($n = 4$)-fold increase of the calnexin signal following 2D-PAGE, while the actin signal was unaffected (Figure 2E, right panels). The increase in calnexin staining upon DHHC6 silencing demonstrated that at least 90% of cellular calnexin is DHHC6 modified at steady state in HeLa cells. Since both sites can be modified (Figures 1B and 2C; Supplementary Figure S2C, E and F), depalmitoylation is slow and must be slow on both sites (Figure 1C), the acylated calnexin population revealed by the 2D gel analysis is most likely modified on both sites.

In Figure 2D, we show that overexpression of DHHC6 leads to a ≈ 2.5 -fold increase in the incorporation of ^3H -palmitate into calnexin. This might at first appear inconsistent with the observation that $>90\%$ of calnexin is palmitoylated. It is however important to note that only calnexin molecules with free cysteines can incorporate ^3H -palmitate. Thus, the bulk of cellular calnexin ($>90\%$) is 'silent' in the ^3H -palmitate incorporation assay and we are monitoring the incorporation of ^3H -palmitate in the remaining population. The rate at which this small population of calnexin with free cysteines becomes modified depends on the abundance of the enzyme. Thus, upon DHHC6 overexpression, incorporation rates increase due to an increased amount of the enzyme and so does the signal after our 2 h standard incorporation (Figure 2D).

Palmitoylation is predicted to affect the conformation of calnexin

Using predictive computational methods, we next analysed whether palmitoylation is expected to modify the structure of the transmembrane domain (TMD) and/or cytosolic tail of calnexin. Using a topology prediction algorithm (Viklund *et al*, 2008), we first predicted that the membrane-embedded residues span from Trp-482 to Cys-502. The TMD was then modelled as an ideal α -helix based on secondary structure predictions and consistent with the UniProt annotations (P27824). We further extended the helical sequence to include residues Glu-478 through Ser-510, in order to gain information on the conformation of the cytosolic tail. The atomistic transmembrane model was inserted into a lipid membrane bilayer and refined using molecular dynamics (MD) simulations under physiological-like conditions (e.g., pH ~ 7 , 1 a.t.m., 298 K). Only small fluctuations of the secondary structure were observed, which support the predicted helical conformation and the match with the hydrophobic interior of the bilayer. The helix tilted by $\sim 30^\circ$ with respect to the membrane surface normal. A distortion of $30 \pm 5^\circ$ in the direction of the helix axis was observed at the level of Pro-494 (Figure 3A), as expected from the helix breaking capacity of this residue (Chang *et al*, 1999). This proline thus breaks the rotational symmetry of the helix with respect to its main axis. The full conservation of this residue in calnexin and calmegins (Supplementary Figure S1) suggests that the TMD kink might be of functional importance.

We next included palmitate moieties at positions Cys-502 (CyP502), Cys-503 (CyP503) or both (CyP502-503). Palmitoylation did not affect the proline-induced kink and only marginally affected the helix tilt. The MD simulations however predicted a palmitoyl-dependent orientation of the cytosolic tail with respect to the helix axis (Figure 3B). Interestingly, palmitoylation at position 503 is predicted to have a more pronounced effect on the conformation of the cytosolic tail, indicating that both sites are not equivalent and that one of the two might have a regulatory role, predictions that will be tested in the future. Combined with the asymmetric nature of the TMD, these simulations raise the interesting possibility that palmitoylation may affect the conformation of calnexin, which in turn could modify its

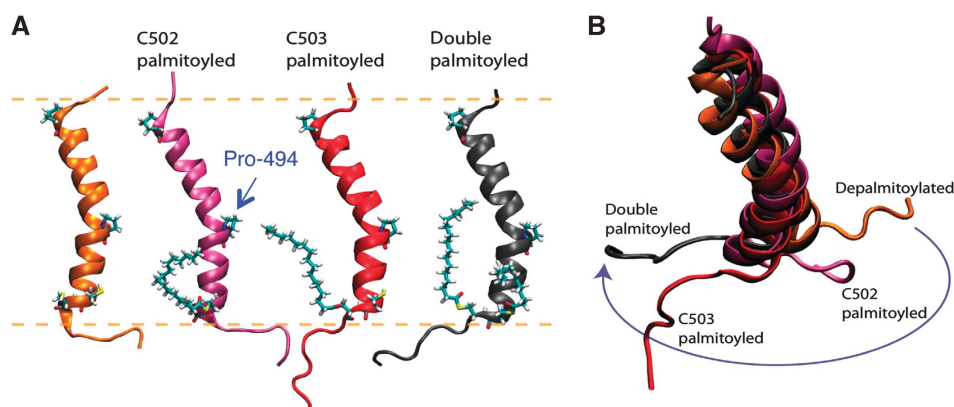


Figure 3 Modelling of the effect of palmitoylation on calnexin conformation. (A) The atomistic models of the helical TMD of calnexin were equilibrated in a solvated DOPC membrane bilayer using MD simulations. Cys-502, Cys-503 and Pro-494 are shown in licorice representation. The approximate position of the membrane bilayer is shown by dashed lines. (B) Palmitoylation affects the orientation of the cytosolic tail of calnexin with respect to the axis of the transmembrane helix.

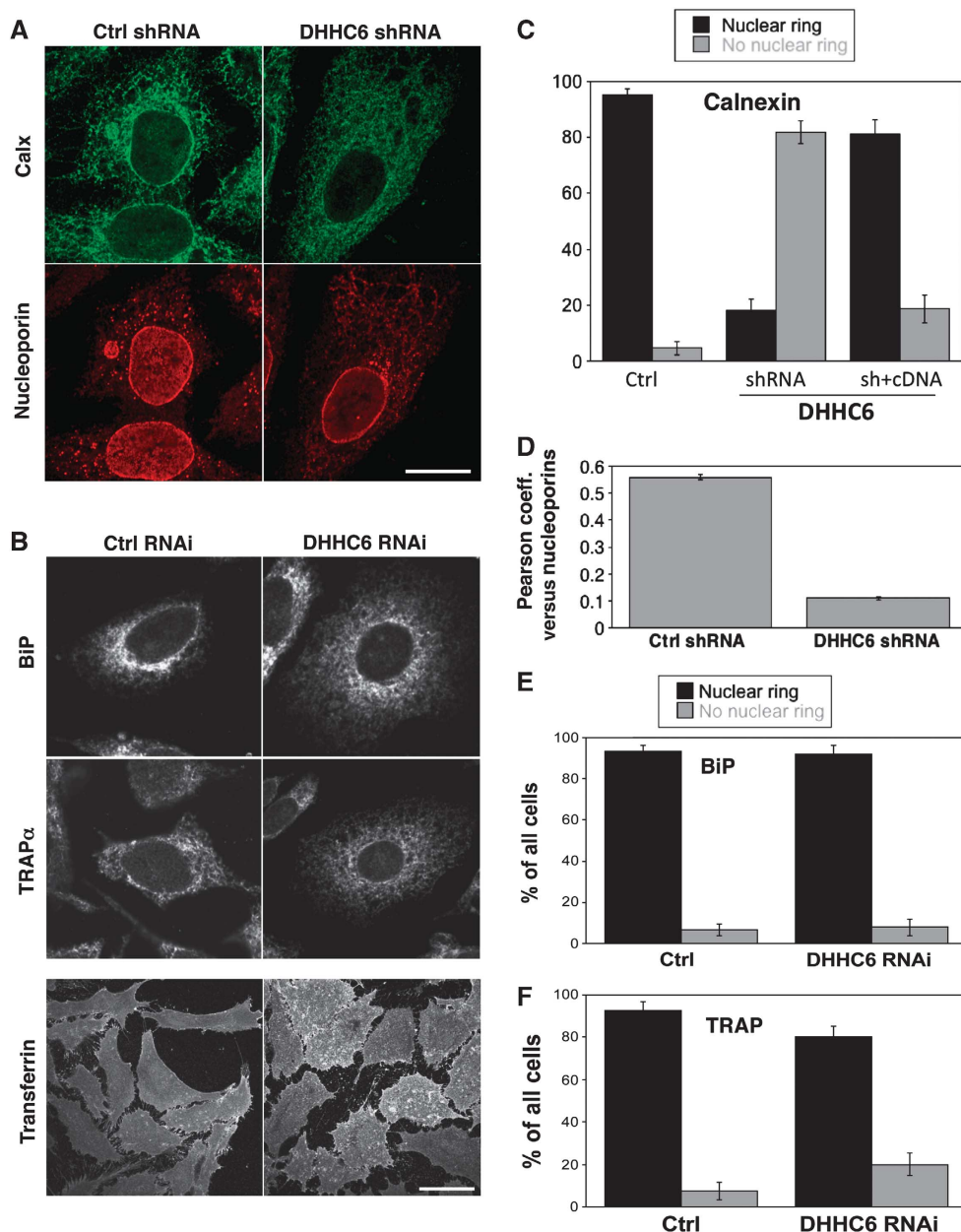


Figure 4 Effect of DHHC6 silencing on calnexin localization. (A) HeLa cells were transfected with either a control shRNA or shRNA against DHHC6 for 8 days, fixed in methanol and stained with anti-calnexin and anti-nucleoporin antibodies. Bar: 10 μ m. (B) HeLa cells grown on cover slips were transfected for 72 h with either control siRNA or siRNA against DHHC6 and were incubated or not with fluorescent Transferrin, washed, fixed in methanol and immunostained with any one of the following antibodies: calnexin, BiP, TRAP α . Each image represents the sum of all the stacks taken in z-axis. To visualize more cells see Supplementary Figure S4A. Bar: 10 μ m. (C) HeLa cells were transfected with either a control shRNA or shRNA targeting DHHC6 for 8 days. On day 6, the cells were transfected with either an empty vector or human DHHC6 cDNA bearing a myc-tag for complementation. The cells were fixed and stained with anti-calnexin, anti-nucleoporin and Hoechst dye. The cells were analysed manually for the presence or absence or restoration of the nuclear membrane staining of calnexin. In all, 50 cells were analysed per condition per experiment. Error bars represent the standard deviation ($n = 3$). (D) The cells from (A) were imaged and the Pearson correlation coefficient for co-localization between calnexin and nucleoporins was determined by monitoring 15 cells/experiment from each condition ($n = 3$). (E, F) Confocal stacks of cells treated as in (B) were analysed manually and the number of cells showing nuclear or no nuclear membrane staining for calnexin (C), BiP (D) or TRAP α was determined. Three independent experiments were analysed and 50 cells were counted for each experiment. Error bars represent standard deviations.

affinity for membrane domains and/or its capacity to interact with proteins in the membrane or in the cytoplasm.

DHHC6 affects calnexin localization

Palmitoylation of membrane proteins has been proposed to modify their affinity for specific plasma membrane micro-

nano domains (Charollais and Van Der Goot, 2009; Levental *et al*, 2010). We therefore investigated whether DHHC6 silencing would alter the distribution of calnexin. Under control conditions, immunofluorescence staining of calnexin is observed in all three major ER compartments (Figure 4A, for additional images see Supplementary Figure S4A),

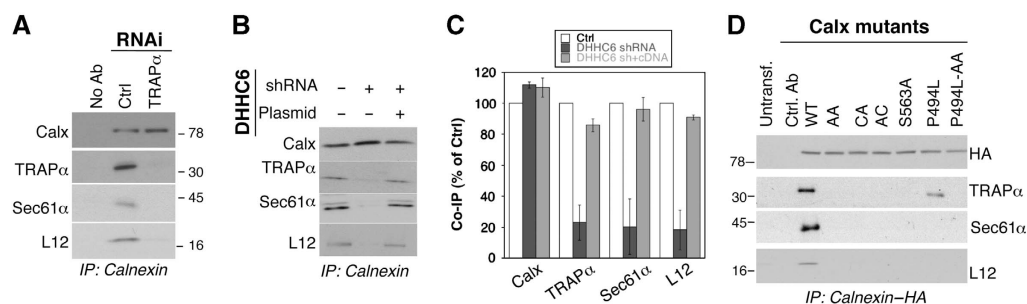


Figure 5 Palmitoylation regulates the interaction of calnexin with the RTC. (A) HeLa cells were transfected for 72 h with either the control siRNA or a siRNA against TRAP α . Immunoprecipitates against anti-calnexin were analysed by western blotting for calnexin, TRAP α , Sec61 α and L12. (B) HeLa cells were transfected for 8 days with control shRNA or the shRNA against DHH6. On day 6, the cells were retransfected either with an empty plasmid (lanes 1 and 2) or with a plasmid-expressing DHH6. The cells were lysed on day 8 and immunoprecipitated using mouse anti-calnexin antibody. Immunoprecipitates were analysed SDS-PAGE followed by western blotting against calnexin, TRAP α , Sec61 α and L12. For the analysis of the total cell lysates see Supplementary Figure S6A. (C) Western blots from (B) were quantified using the Image J software and the error bars represent the standard deviation ($n = 3$). (D) HeLa cells were transfected for 48 h with either WT, single or double cysteine calnexin mutants, mutant of the phosphorylation calnexin site S563A or of the TMD proline. Immunoprecipitates against anti-HA were analysed by western blotting against: TRAP α , Sec61 α , Sec62, Sec61 β and L12. For the analysis of the total cell lysates see Supplementary Figure S6B.

resembling the staining of the soluble luminal ER chaperone BiP and the transmembrane RTC-associated protein TRAP α (Figure 4B, for additional images see Supplementary Figure S4A).

When silencing DHH6, the first observation was a change in cell morphology: cells were somewhat less elongated (see Transferrin staining bottom of Figure 4B) and the average cell footprint was $\sim 30\%$ larger (Supplementary Figure S5A). The nuclear staining and size was however unaltered (Figure 4A). When viewing through confocal stacks of a lawn of cells stained for calnexin (Supplementary Movie S1A versus Supplementary Movie S1B), it readily appeared that calnexin staining was lost from the nuclear envelope upon DHH6 silencing. This was confirmed by quantifying the cells showing a nuclear envelope calnexin staining and those that did not, by eye for multiple cells over multiple experiments (Figure 4C). In agreement, the Pearson coefficient of co-localization of calnexin with nucleoporins strongly decreased upon DHH6 silencing (Figure 4D; Supplementary Figure S4B). Importantly, nuclear membrane localization of calnexin was restored upon recombination with the DHH6-expressing plasmid (Figure 4C; Supplementary Figure S4B). Disappearance from the nuclear membrane was specific to calnexin, since BiP and TRAP α staining was unaffected by DHH6 silencing (Figure 4B, E and F; Supplementary Figure S4A). While disappearance from the nuclear membrane was obvious, other changes in calnexin staining, in particular in the dense and complex perinuclear ER could not be quantified in satisfactory manner. Our interpretation of the change in calnexin distribution is that palmitoylation increases the affinity of calnexin for ER sheets, which encompasses the entire rough ER and of which the nuclear membrane is the most exacerbated form.

Palmitoylation regulates the interaction of calnexin with the RTC

The observation that DHH6 expression favours the localization of calnexin to the rough ER prompted us to investigate whether calnexin interacts with the RTC under steady-state conditions. The central translocon component sec61 α , as well

as the translocon-associated protein TRAP α and the ribosomal subunit L12 were recovered upon immunoprecipitation of endogenous calnexin (Figure 5A). The interaction of calnexin with sec61 α and L12 was abolished upon RNAi silencing of TRAP α (Figure 5A), indicating first that the observed interaction of calnexin with sec61 α and L12 is specific, and not due to bulk co-isolation of membrane proteins in detergent/lipid micelles. Moreover, the observation indicates that the TRAP complex mediates the interaction of calnexin with the RTC. Interestingly, TRAP α silencing did not affect the localization of calnexin to the nuclear membrane (Supplementary Figure S5B), showing that DHH6-induced perinuclear calnexin localization is not due to its interaction with RTCs.

We next silenced DHH6 by shRNA treatment for 8 days, in order to reduce the cellular pool of S-acylated calnexin. Upon long-term DHH6 silencing, co-immunoprecipitation of TRAP α , sec61 α and L12 with calnexin decreased by $\approx 80\%$ (Figure 5B and C; Supplementary Figure S6A). Importantly, interaction of calnexin with the RTC was restored by transfection with DHH6-expressing plasmid (Figure 5B and C; Supplementary Figure S6A). RTC interaction experiments were next carried out using the single and double cysteine mutants of calnexin, the S563A mutant, where the phosphorylation site promoting ribosome association is modified (Chevet *et al*, 1999), and the P494L mutant, where the transmembrane proline (Figure 3A) is substituted with a hydrophobic residue to remove the kink in the TM helix. It should be noted that the two latter mutations did not prevent palmitoylation (personal communication). The calnexin-RTC interaction was abolished upon mutation of Ser-563 to alanine (Figure 5D; Supplementary Figure S6B), indicating the importance of this residue even in resting cells (Chevet *et al*, 1999). The RTC interaction was also severely affected by the P494L mutation, pointing out a central role of TMD conformation for the association (Figure 5D). Most strikingly, mutation of either S-acylation site abolished the calnexin-RTC interaction (Figure 5D). Thus, not only must calnexin be acylated to interact with the RTC, it has to be modified on both cysteines.

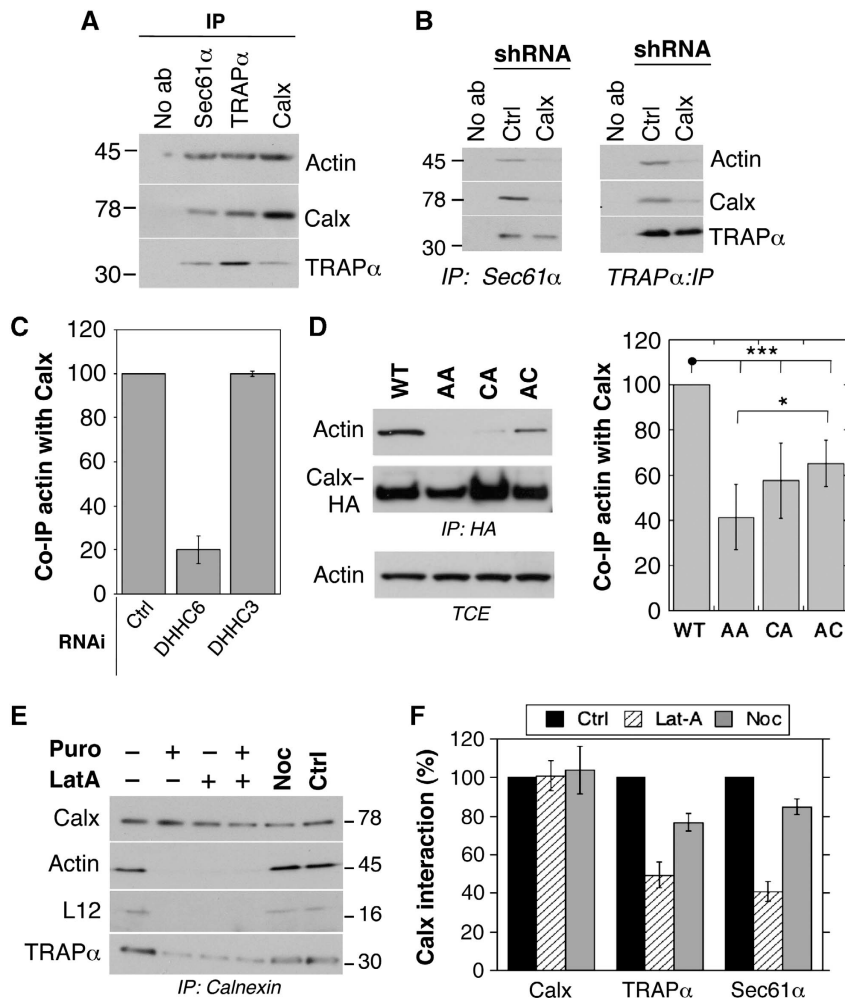


Figure 6 Calnexin palmitoylation promotes interaction with actin. (A) HeLa cells lysates were immunoprecipitated using either anti-Sec61α anti-TRAPα, anti-calnexin antibodies or no antibody. The immunoprecipitates were analysed by western blotting against actin, calnexin and TRAPα. Blotting against Sec61α could not be performed as it migrates at the same molecular weight as the antibody heavy chain. (B) HeLa cells were transfected for 6 days with either shRNA against luciferase or shRNA against calnexin and lysed. After immunoprecipitation against Sec61α (left panel) or TRAPα (right panel), samples were analysed by western blotting against actin, calnexin and TRAPα. As in (A), Sec61α could not be probed. (C) HeLa cells were transfected with a control siRNA, siRNA against DHHC6 or DHHC3 and lysed. Immunoprecipitates against calnexin were analysed by western blotting against calnexin and actin. Quantification was performed using the Typhoon Imager. Error bars correspond to standard deviations ($n = 4$). (D) HeLa cells were transfected or not for 48 h with cDNA-expressing calnexin-HA-WT, double or single cysteine mutants, and lysed. Immunoprecipitates against HA were analysed by western blotting against HA and actin and quantified using the Typhoon Imager. Error bars correspond to standard deviations ($n = 4$). Total cell extracts (TCE) were probed for actin. (E) HeLa cells were treated for 30 min at 37°C with either DMSO, Latrunculin A or Nocodazole or 15 min with puromycin or water, which was used as a control. The cells were lysed and immunoprecipitated with rabbit anti-calnexin antibody. The immunoprecipitates and the total cell extracts were migrated on SDS-PAGE followed by western blot to reveal calnexin, actin, TRAPα and L12 ($n = 3$). (F) The treatment with Latrunculin A and Nocodazole was performed as in (E). Immunoprecipitates against calnexin were analysed by SDS-PAGE and followed by western blotting against calnexin, TRAPα and Sec61α. Western blots were quantified using the Image J software and the error bars correspond to standard deviations ($n = 3$). *** $P < 0.001$, * $P < 0.05$.

Actin is recruited and stabilizes the calnexin-RTC supercomplex

Stabilization of certain membrane protein complexes/clusters has been found to be dependent on the actin cytoskeleton (Abrami *et al*, 2010). We probed whether the calnexin-RTC supercomplex interacts with actin. Actin was pulled down upon immunoprecipitation of sec61α, TRAPα as well as calnexin (Figure 6A). The interaction between sec61α or TRAPα with actin was strongly diminished upon silencing of calnexin (Figure 6B), indicating first that the interaction observed in Figure 6A is specific and not due to contamination by this abundant protein, and second that calnexin plays a crucial role in the binding of actin to the RTC.

The interaction of calnexin with actin was dependent on its palmitoylation status as indicated by the decrease in actin association upon DHHC6 silencing (Figure 6C). Calnexin-actin interaction was restored upon transfection with the DHHC6-expressing plasmid (Supplementary Figure S6C). The importance of palmitoylation for calnexin-actin interaction was confirmed by the weak association of the single and double cysteine calnexin mutants with actin (Figure 6D). The fact that DHHC6 silencing had a more pronounced effect than mutation of the palmitoylation sites might be due to indirect effects of unidentified ER DHHC6 targets.

Altogether, the above observations raised the possibility that actin can only be recruited to calnexin-RTC supercomplex

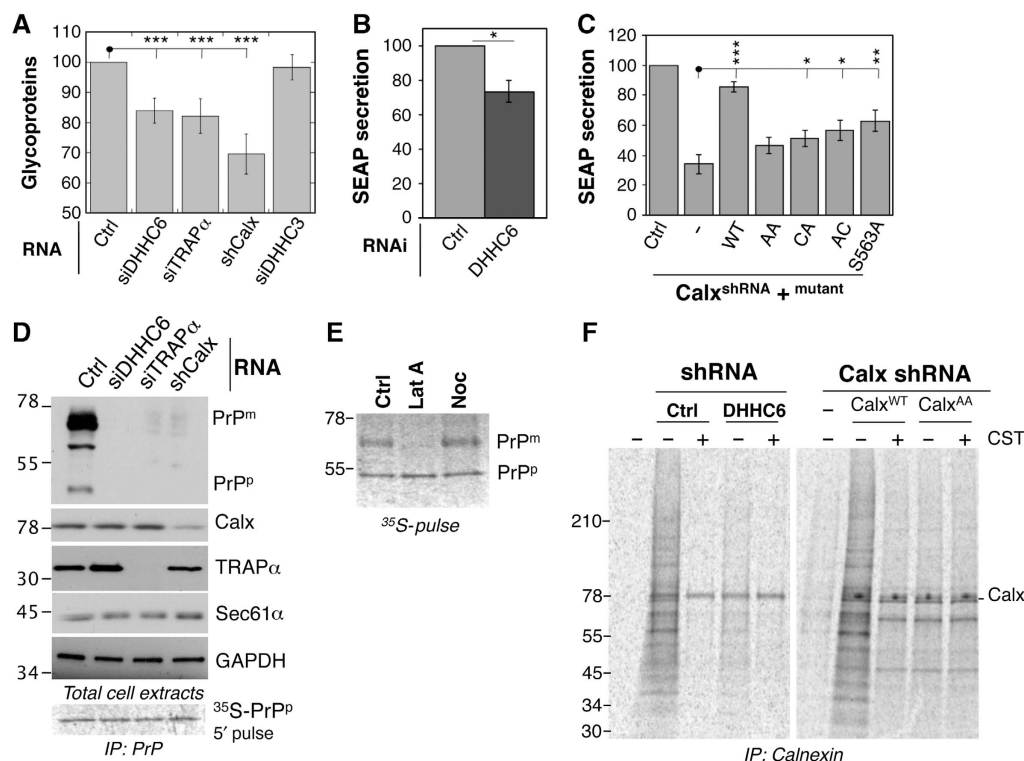


Figure 7 Palmitoylation of calnexin promotes the folding of glycosylated proteins. **(A)** HeLa cells were transfected with either the control siRNA or siRNAs against DHHC6, TRAP α or DHHC3 for 72 h and shRNA against calnexin for 144 h. The cells were labelled with 35 S-methionine/cysteine for 20 min. The cytosolic proteins were isolated by semi-permeabilizing the cells using digitonin. The glycoproteins were separated from the non-cytosolic fraction by using Con-A agarose beads. The glycoprotein fraction was run on SDS-PAGE and the radiolabelled products were visualized by Typhoon phosphorimager. The radiolabelled lanes were quantified using the Typhoon Imager and the histograms were plotted as the ratio of glycoproteins to the cytosolic proteins in each sample. The error bars represent the standard deviation ($n=4$). *** $P<0.001$. For the autoradiograms of glycoproteins and cytosolic proteins see Supplementary Figure S7A and B. **(B)** HeLa cells were co-transfected with either control siRNA or siRNA against DHHC6 and the pSEAP2 plasmid. The cell medium was replaced with fresh medium in all the plates so that the SEAP secreted in the final 24 h was quantified. The SEAP secreted was normalized to the total amount of protein present in each sample and expressed as percentage of SEAP secreted by the control. * $P<0.05$. **(C)** HeLa cells were transfected with either the control shRNA or the shRNA against calnexin. At 96 h post-transfection, cells were transfected with the pSEAP2 vector and either empty vector or the vector-expressing WT or mutant calnexins. At 120 h post-transfection, the cell medium was replaced with fresh medium in order to assess the SEAP accumulated in the last 24 h. The measured SEAP values were normalized to the total amount of the protein present in each sample and the values were plotted as a percentage of the shLuc transfection. *** $P<0.001$, ** $P<0.01$, * $P<0.05$. **(D)** HeLa cells were either co-transfected with GFP-PrP and the control siRNA or the siRNA targeting DHHC6, TRAP α . Calnexin was silenced by shRNA and GFP-PrP was transfected 96 h after the shRNA transfection. Cell lysates were prepared and equal amounts of the proteins were analysed by western blotting for GFP, calnexin, TRAP α , Sec61 α , GAPDH. PrP^p indicates the precursor form of GFP-PrP and PrP^m the mature form. **(E)** HeLa cells were transfected with GFP-PrPwt for 24 h, treated with either DMSO, Latrunculin A or Nocodazole for 30 min at 37°C and prior to 30 min 35 S-methionine-cysteine pulse labelling in the presence of the drugs. The cell lysates were immunoprecipitated with anti-GFP antibody and the immunoprecipitates were analysed by SDS-PAGE followed by Typhoon phosphorimaging. PrP^p indicates the precursor and PrP^m the mature form of the GFP-PrP protein. **(F)** To monitor the binding of calnexin to the substrates, HeLa cells were treated with either shRNA against DHHC6 (left panel) or shRNA against calnexin (right panel). The cells depleted of endogenous calnexin were complemented with either the WT or AA mutant. The cells were pulse labelled with 35 S-methionine-cysteine for 10 min followed by a chase of 3 min either in the presence of CST or not. The lysate was immunoprecipitated using anti-calnexin antibody and migrated on a SDS-PAGE followed by fixation and drying of the cells. The radiolabelled products were revealed by Typhoon phosphorimager.

when it is fully assembled. To test this possibility, we stripped off ribosomes from the ER membrane by a 15-min puromycin treatment. Strikingly, calnexin then no longer interacted with either the translocon or actin (Figure 6E), indicating that indeed actin can only bind to the fully assembled supercomplex. Moreover, the puromycin effect shows that calnexin is recruited to the translocon only when the ribosome has docked.

In reverse, we investigated whether actin polymerization was important for the stabilization of the supercomplex. Inhibiting actin polymerization by a short latrunculin A treatment was sufficient to decrease the interaction of calnexin with L12, TRAP α and sec61 α (Figure 6E and F), while nocodazole, which disrupts microtubules, had no effect (Figure 6E and F).

Palmitoylation of calnexin promotes the folding of glycosylated proteins

Finally we investigated whether the calnexin-RTC interaction is important for the lectin chaperone function. We first analysed the effect of silencing either calnexin, TRAP α or DHHC6 on the production—that is, an overall read out for synthesis, folding and secretion—of total glycoproteins. Silencing of any of the three proteins led to a significant decrease in total glycoproteins as monitored by a 20-min 35 S Cys/Met pulse, followed by precipitation with the lectin, concavalin A (Figure 7A; Supplementary Figure S7A). The 30% drop in detected glycoproteins is consistent with the known redundancy of the folding systems for glycoproteins (Braakman and Bulleid, 2011). As a control, we monitored the

production of cytosolic, that is, non-*N*-glycosylated, proteins and found that silencing of calnexin, TRAP α or DHHC6 had no effect (Supplementary Figure S7B).

Next, we analysed the production of two specific glycoproteins: a soluble version of alkaline phosphatase (SEAP, secreted alkaline phosphatase), which provides a quantitative readout upon secretion, again used as an overall read out for synthesis, folding and secretion (Lakkaraju *et al*, 2008) and the prion protein (PrP), which undergoes a characteristic change in migration by SDS-PAGE upon folding and maturation (Fons *et al*, 2003).

DHHC6 silencing led to a 30% decrease in SEAP secretion (Figure 7B). In contrast, secretion of the non-glycosylated protein bovine preprolactin, a well-studied protein translocation substrate (Lakkaraju *et al*, 2008), was not affected by DHHC6 silencing (Supplementary Figure S7C).

To assess the direct involvement of calnexin and the importance of its palmitoylation status, we silenced calnexin by shRNA and reconstituted the cells with wild-type (WT) or mutant calnexins (Supplementary Figure S7D). Calnexin silencing led to a 70% drop in SEAP secretion, which could be restored to >80% of control cells, upon reconstitution with the WT protein (Figure 7C). In contrast, no recovery was observed upon reconstitution with palmitoylation-deficient calnexin (AA mutant) (Figure 7C), while single cysteine and S563A mutants led to a very weak recovery (Figure 7C). Altogether, these observations show that DHHC6-mediated palmitoylation of calnexin on both cysteines is required for optimal folding and secretion of SEAP.

We next monitored maturation of GFP-tagged PrP, which upon expression in control cells migrates by SDS-PAGE predominantly in a mature form with some precursor form (Figure 7D). Neither the mature nor the precursor PrP forms were detectable by western blotting upon silencing of calnexin, TRAP α or DHHC6. Synthesis of PrP precursor was, however, unaffected by the calnexin, TRAP α or DHHC6 silencing as determined by a 5-min ³⁵S-pulse labelling of PrP (bottom of panel Figure 7D and quantification Supplementary Figure S7E). Thus, in the absence of these proteins, PrP is rapidly degraded following synthesis. A similar rapid post-translational degradation of PrP was observed upon Latrunculin A treatment (Figure 7E), further supporting that proper folding of PrP requires the fully assembled calnexin-RTC supercomplex.

Altogether, the above experiments indicate that *S*-acylation controls the ability of calnexin to act as a chaperone. To address this point specifically, we monitored the binding of monoglucosylated substrates to calnexin. DHHC6 was silenced or not using shRNA and newly synthesized proteins were metabolically labelled. After detergent solubilizing the cells, calnexin-substrate complexes were isolated by immunoprecipitation against calnexin. As a control, cells were pretreated or not with castanospermine (CST), a specific inhibitor of α -glucosidases, which prevents the generation of monoglucosylated proteins (Hammond and Helenius, 1994). Importantly, CST had no effect on either palmitoylation or depalmitoylation of calnexin (not shown). Also, castanospermine did not alter the cellular localization of calnexin nor its interaction with the RTC (not shown).

After a pulse of metabolic labelling, a well-defined pattern of calnexin substrates was detected in control cells, which was absent in CST-treated cells (Figure 7F, left panel). DHHC6

silencing led to a drastic decrease in the amount of calnexin-bound substrates, without affecting their pattern (Figure 7F, left panel). Even when the metabolic pulse was followed by a chase period, substrate binding could not be observed in DHHC6-silenced cells, indicating that not only co- but also post-translational binding was defective. To confirm that non-acylated calnexin is unable to capture its substrates, substrate binding of the AA palmitoylation-deficient mutant calnexin was monitored in comparison to WT, in cells where endogenous calnexin was silenced by shRNA. As shown in Figure 7F (right panel), palmitoylation-deficient calnexin failed to bind its substrates. This observation, however, does not necessarily mean that the capacity of calnexin to binding monoglucosylated proteins *per se* is affected. It reveals that in the cellular context, calnexin must be palmitoylated and be part of the RTC complex to grab its substrates. If substrates are not captured co-translationally, they either aggregate or are handled by other folding systems in the ER, but calnexin will not bind them post-translationally.

Discussion

Calnexin is a transmembrane chaperone that aids in the folding of glycosylated proteins by binding to monoglucose residues on the branched N-linked oligosaccharide and thus protects the protein from aggregation. The interaction of calnexin with nascent chains has been observed >15 years ago (Chen *et al*, 1995), but little information was available on the specific mechanisms that may facilitate this co-translational interaction. We here report that calnexin is a stable component of the RTCs, as is the OST, leading to the formation of a supercomplex. This interaction occurs via the translocon-associated complex TRAP. Based on X-ray crystallography and single particle analysis, a model of the translocon complex has been proposed wherein TRAP α interacts with sec61 α , the latter interacting via the opposing face of the molecule with sec61 β (Menetret *et al*, 2008). Our findings thus allow a rough positioning of calnexin in this supercomplex, sandwiching TRAP α between calnexin and the translocon (Figure 8).

Formation and stabilization of calnexin-RTC supercomplex requires several factors. As previously described (Chevet *et al*, 2010), Ser-563 is critical possibly due to its phosphorylation. The kinked conformation of the TMD also influences the interaction, as indicated by the disruptive effect of mutating the highly conserved TM Pro-494. Importantly, calnexin must be palmitoylated.

Despite the presence of 12 (Ohno *et al*, 2006) to 16 (our unpublished observations in HeLa cells) palmitoyltransferase in the mammalian ER, the *S*-acylation of calnexin is performed by a single enzyme, DHHC6. The involvement of a single DHHC enzyme was rather unexpected since these enzymes are thought to have broad overlapping specificities at least for soluble substrates (Greaves and Chamberlain, 2011). Also, in contrast to soluble substrates such as Ras, calnexin does not undergo dynamic cycles of palmitoylation-depalmitoylation. As a consequence, at steady state, the *S*-acylated form, and most likely the dual acylated form, accumulates and almost the entire pool of cellular calnexin is modified in resting cells as shown by our 2D gel analysis. The observed slow rate of palmitate turnover might be due to the close proximity, and possibly even embedding, of the cysteine

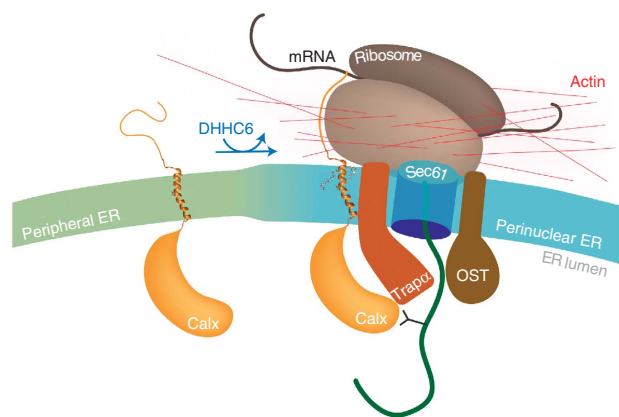


Figure 8 Molecular consequences of calnexin palmitoylation. When its palmitoylation sites are free, calnexin preferentially localizes to the peripheral tubular ER. Upon S-acylation by DHHC6, calnexin partitions preferentially into the rough ER and associates with TRAP α , altogether ensuring that calnexin efficiently interacts with the RTC. Once the calnexin–RTC has assembled into a supercomplex, actin is recruited and stabilizes the assembly. In this configuration, calnexin is positioned to efficiently grab nascent glycoproteins as they emerge from the translocon pore, acquire their N-linked glycan from which the two external glucoses have been trimmed.

residues in the membrane, which would render them inaccessible to cytosolic thioesterases, of which only one has so far been identified, the soluble enzyme APT1 (Dekker *et al*, 2010). Recently, it has been observed for β 1-adrenergic receptor that the susceptibility of cysteines to thioesterases indeed correlates with distance from the TMD (Zuckerman *et al*, 2011).

S-acylation of calnexin promotes its interaction with the RTC in two ways: First, palmitoylation affects the ability of calnexin to interact with the TRAP complex, possibly by influencing the conformation of the TMD and/or the cytosolic tail as suggested by the MD simulations. Second, palmitoylation favours the association of calnexin with the nuclear membrane, and probably with sheet-like ER structures in general. Calnexin spans the membrane only once and it is therefore unlikely that it has any membrane curvature sensing capacity. Since protein and lipid compositions are likely different in ER tubules versus ER sheets (Shibata *et al*, 2008, 2010), palmitoylation may favour the partitioning of calnexin to sheet-like structures, in a manner similar to raft partitioning of transmembrane proteins at the plasma membrane (Levental *et al*, 2010). Calnexin has also been shown to localize to mitochondrial-associated membranes (MAM), sites of interaction between the ER and mitochondria (Myhill *et al*, 2008; de Brito and Scorrano, 2010). During the revision of this manuscript, localization of calnexin as well as of TMX, a transmembrane ER-thioredoxin, to MAMs was reported to depend on palmitoylation of these two type I membrane proteins (Lynes *et al*, 2011). Future studies should elucidate whether and how S-acylation-dependent localization of calnexin to the various ER domains is controlled.

Interestingly, once fully assembled, the calnexin–RTC supercomplex recruits the actin cytoskeleton, which in turn is required to stabilize the assembly. While in yeast and plants where cortical ER is known to move along actin cables (Du *et al*, 2004), few reports describe interactions between actin and the ER in animal cells (Du *et al*, 2004). Actin is well known to interact with the nuclear envelope via Nesprins,

large transmembrane proteins involved in nuclear positioning and nuclear envelope architecture (Lombardi *et al*, 2011). However, interactions with the rest of the ER have not been reported with the exception of a study showing that actin controls the diffusion of newly synthesized glycoproteins in the ER lumen and thereby reduces aggregation during folding (Nagaya *et al*, 2008). Our observations raise the possibility that this is due to the handling of these glycoproteins by the calnexin–RTC supercomplexes.

Altogether, the present work leads to a model where DHHC6-mediated S-acylation favours the partitioning of calnexin into the rough ER and, in addition, allows it to interact with RTC components leading to the formation of a supercomplex, which subsequently recruits the actin cytoskeleton, the polymerization of which in turn is required to maintain the assembly (Figure 8). When the nascent polypeptide chain emerges from the translocon pore, it arrives into a folding prone microenvironment: the OST and BiP initially keep it in an unfolded state to allow optimal N-glycosylation and also await the full synthesis of folding units. By being an integral part of the ribosome–translocon supercomplex, calnexin is ideally positioned to capture its client protein following appropriate glucose trimming, protecting the protein from aggregation as it undergoes oxidative folding, further benefiting from the calnexin-associated oxidoreductase ERp57.

Materials and methods

Cell culture, plasmids, transfections and real-time PCR

HeLa cells (ATCC) were grown in complete modified Eagle's medium (MEM) (Sigma) at 37°C supplemented with 10% fetal bovine serum (FBS), 2 mM L-glutamine, penicillin and streptomycin. RPE1 cells were grown in complete Dulbeccos MEM (DMEM) (Gibco) supplemented with 10% FBS, 2 mM L-glutamine, penicillin and streptomycin. Mammalian expression plasmids harbouring human C-terminally tagged Calnexin–HA and Calnexin–GFP were generated by cloning the calnexin cDNA into pCDNA3 vector bearing the HA tag or GFP. Constructs to express single- or double-point mutation of calnexin were generated using the Quickchange (Stratagene) strategy following the manufacturer's instructions. Isoform 1 of human tumour endothelial marker 8 (TEM8) cDNA tagged with a HA epitope was cloned in the pIRESHyg2 vector. Plasmid encoding human myc–LRP6 was provided by Dr G Davidson. GFP–PrP was provided by Dr Chiara Zurzolo. PPL-3f was provided by Dr Katharina Strub. Human DHHC cDNA were cloned in pCE puro–His–myc or flag-expressing plasmids. Plasmids were transfected into cells for 24 or 48 h (2 μ g cDNA/9.6 cm² plate) using Fugene (Roche Diagnostics Corporation). For control transfections, we used an empty pCDNA3 plasmid. shRNA against calnexin were generated from the p.Super.Retro.Puro vector containing the hairpin insert against the following target sequence present at the 3' UTR of calnexin gene: 5'-gagcttgatctgtgatttc-3' (Supplementary Table S1). As a control, we used shRNA against the target sequence in firefly luciferase gene: 5'-CGTACGCGGAATACTT CGA-3'. shRNA against DHHC6 was generated by cloning the hairpin insert against the target region in the 3' UTR of DHHC6: 5'-CCTAGTGGCATGATTAAAG-3' (Supplementary Table S1).

For the real-time PCR, RNA was extracted from a six-well dish using the RNeasy kit (Qiagen). In all, 1 mg of the total RNA extracted was used for the reverse transcription using random hexamers and superscript II (Invitrogen). A 1:40 dilution of the cDNA was used to perform the real-time PCR using SyBr green reagent (Roche). mRNA levels were normalized using three housekeeping genes: TATA-binding protein, β -microglobulin and β -glucuronidase. Total RNA of different human tissues was obtained from Amsbio.

Antibodies and reagents

Polyclonal antibodies against calnexin were generated against the C-terminal peptide: CDAEEDGGTVSQEEDRKPK in rabbit; rat

antibodies against purified CMG2 were produced in our laboratory; anti-HA and anti-GFP were from Roche; Anti-flag M2 antibody from Sigma; anti-HA-agarose-conjugated beads, used for the immunoprecipitations from Roche (Applied Science, IN); anti-calnexin monoclonal and anti-myc antibodies from Santa Cruz; anti-actin antibodies from Millipore; anti-Transferrin receptor antibodies from Zymed; anti-Nuclear Pore Complex Proteins (Mab414), anti-Sec61 α , anti-BiP antibodies from Abcam. Anti-L12 antibodies were provided Dr K Strub. Anti-TRAP α antibody was provided by Dr R Hegde. Protein G-agarose-conjugated beads were from GE Healthcare, HRP secondary antibodies were from Pierce, and Alexa-conjugated secondary antibodies from Molecular Probes. Latrunculin A (used at a concentration of 0.4 mg/ml for 30 min or 1 h) and Nocodazole (used at a concentration of 10 μ M for 30 min or 1 h) were from Sigma. CST was purchased from Calbiochem and used at a concentration of 1 mM. Puromycin was purchased from Calbiochem and used at a concentration of 200 μ M for 15 min to strip the ER of ribosomes in HeLa cells.

Immunoprecipitation and RNAi experiments

siRNA against human DHHC were purchased from Qiagen (see Supplementary Table S1). As control siRNA we used the following target sequence of the viral glycoprotein VSV-G: 5'atgaacaaacgaaacaagga3'. For gene silencing, HeLa cells were transfected for 72 h with 100 pmol/9.2 cm² dish of siRNA using interferrin (Polyplus) transfection reagent.

For all immunoprecipitations unless specified, cells were lysed 30 min at 4°C in IP buffer (0.5% NP-40, 500 mM Tris-HCl pH 7.4, 20 mM EDTA, 10 mM NaF, 2 mM benzamidine, and a cocktail of protease inhibitors; Roche) followed by centrifugation for 3 min at 2000 g. The supernatants were precleared with protein G-agarose-conjugated beads and incubated 16 h at 4°C with antibodies and beads. The beads were washed for three times with the IP buffer and resuspended in the sample buffer (2 \times) after the final wash. The samples were heated at 95°C for 5 min and migrated on SDS-PAGE. Western blotting was performed using the iBlot (Invitrogen) according to the manufacturer's instructions. Quantification of the blots was done using either the Typhoon Imager or Image J software.

Radiolabelling experiments

To monitor palmitoylation, HeLa cells transfected with either the WT or mutant calnexin constructs were incubated for 2 h at 37°C in IM (Glasgow minimal essential medium buffered with 10 mM Hepes, pH 7.4) with 200 μ Ci/ml ³H palmitic acid (9,10-³H(N)) (American Radiolabeled Chemicals, Inc.). The cells were washed and the cell lysate was extracted followed by immunoprecipitation with tag specific antibodies in the case of exogenously expressed calnexin and calnexin antibody in case of endogenous protein. After the washes, beads were incubated for 5 min at 90°C in non-reducing sample buffer (2 \times) prior to SDS-PAGE. After the SDS-PAGE, the gel was incubated with a fixative solution (25% isopropanol, 65% H₂O, 10% acetic acid), followed by a 30-min incubation with signal enhancer Amplify NAMP100 (Amersham). The dried gels were exposed to a Hyperfilm MP (Amersham). Chemical removal of S-palmitoylation was performed by treating cell extracts for 1 h at room temperature with 1 M hydroxylamine hydrochloride (Sigma) pH 7.2. Protein synthesis was blocked by 1 h treatment with 10 μ M cycloheximide (Sigma) at 37°C.

To monitor the ability of calnexin to bind substrates, HeLa cells were first transfected with the shRNA against calnexin for 7–9 days. After 96 h, cells were retransfected with either the WT or the AA mutant calnexin for 48 h.

Cells were pretreated or not for 1 h with 1 mM CST, a specific inhibitor of α -glucosidases and pulse labelled with 100 μ Ci/ml ³⁵S-methionine/cysteine for 10 min followed by a 3-min chase in the presence or absence of drug. Cells were harvested in an isotonic HEPES buffer (pH 6.8) containing 2% CHAPS and protease inhibitor cocktail. Post-nuclear supernatants, obtained by centrifuging the sample at 10 000 g for 10 min, were submitted to immunoprecipitation overnight with anti-calnexin antibody followed by incubation with protein G agarose beads for 2 h at 4°C.

To monitor the synthesis and maturation of GFP-PrP protein, HeLa cells transfected with GFP-PrP under different experimental conditions were labelled with 100 μ Ci/ml ³⁵S-methionine/cysteine for 30 min. The cells were then lysed in lysis buffer composed of 0.1 M Tris-HCl pH 8 and 1% SDS. The cell lysate was heated at 95°C

for 10 min with occasional stirring until the lysate is no more viscous. The cleared lysate was diluted in the RIPA buffer (150 mM NaCl, 1% NP-40, 0.5% deoxycholate, 0.1% SDS, 50 mM Tris-HCl (pH 8), 1 mM EDTA, and 1 \times protease inhibitor cocktail). Immunoprecipitation was performed using an anti-GFP antibody.

For all the experiments, samples were analysed by 4–12% gradient SDS-PAGE, followed by fixing and drying of the gels. The radiolabelled products were revealed using Typhoon phosphorimager and the quantified using the Typhoon Imager (Image QuantTool, GE Healthcare).

Glycoprotein analysis

Fractionation of newly synthesized glycoproteins was performed in 35 mm dishes. Cells were metabolically labelled with 50 μ Ci/ml of ³⁵S-Methionine/cysteine mix for 20 min, followed by a wash with ice-cold PBS. Cytosolic proteins were extracted by treating the cells with 150 μ g/ml of digitonin in KHM buffer (110 mM KAc, 20 mM Hepes, pH 7.2, 2 mM MgAc₂) for 5 min. After the recovery of cytosolic extract, the cells were washed once again in the KHM buffer and resuspended in 500 μ l of the IP buffer to extract the non-cytosolic proteins. The glycoproteins were separated from the non-cytosolic protein fraction by incubation with Con-A beads (GE Healthcare) for 1 h. Equal amounts of total protein were loaded on the Con-A beads from all the samples Con-A beads selectively bind to the glycoproteins. The beads were washed for three times with the IP buffer and the glycoproteins were eluted by 300 μ l of IP buffer containing 0.25 M α -methyl-D-mannopyranoside. Both the cytosolic and glycoprotein fractions were migrated on 4–20% SDS-PAGE gradient gels. The gels were fixed, dried and further analysed for autoradiography.

Calnexin complementation and SEAP assay

For the complementation assay, HeLa cells were transfected with shRNA against calnexin and the transfected cells were selected by treating with puromycin (3 μ g/ml) for 24 h. At 72 h post-transfection, the cells were split into six-well plates and the day after were transfected with the control plasmid, the WT or the mutant calnexin cDNAs and when required with the pSEAP2 reporter plasmid.

SEAP assay was performed using Great EscAPE SEAP Chemiluminescence Kit 2.0 (Clontech). The cell medium was changed 24 h before the SEAP assay was done in order to monitor the SEAP secreted for 24 h when the RNAi or the over expression is most effective. Assay was performed in a 96-well plate, using 15 μ l of the growth medium according to the manufacturer's instructions. The Chemiluminescence signal was collected by Spectra Max multiwell plate reader and the data were analysed by Soft Max Pro 5 software.

DHHC6 recomplementation

For the complementation assay, HeLa cells were transfected with shRNA against DHHC6 followed by selection of transfected cells by treating with puromycin (3 μ g/ml) for 24 h. The cells were further transfected on day 6 with the cDNA-expressing human DHHC6 without its 3' UTR for complementation. The functional assays were performed on day 8 to analyse the restoration of phenotypes observed by the loss of DHHC6.

Immunofluorescence microscopy

Immunofluorescence in HeLa was done as described previously with the exception of the fixative used. To label the plasma membrane of the cells, the cells were treated with 5 μ g/ml of FITC-labelled Transferrin (Molecular Probes) for 1 h at 4°C. This was followed by washes at 4°C and fixation. Cells were fixed and permeabilized with methanol at –20°C for 4 min. Cells were further labelled with the appropriate primary antibody, followed by labelling with Alexa-conjugated goat anti-rabbit or anti-mouse IgG (405, 488, 568 nm). The nuclei were stained by Hoechst dye. Images were acquired using a \times 63 lens on LSM-710 Laser scanning microscope (Carl Zeiss Microimaging, Inc.). The Zen software was used for the processing of the images and the 3D stacks. The movie animations and the summation of the z-stacks were done using the Image J software. To calculate the Pearson correlation coefficient, four different regions on the nuclear membrane were analysed for 15 cells. Each condition was subjected to a total of 60 regions and the Pearson coefficient was calculated using Coloc 2 plugin in FIJI software.

MD simulations

We used MD simulations to characterize the calnexin TM domain at the atomistic level in the membrane environment. As a first step, the ideal helical TM model was inserted and equilibrated in a $60 \times 60 \text{ \AA}^2$ palmitoyl oleoyl phosphatidyl choline membrane patch (Humphrey *et al*, 1996) to characterize its structure and dynamics in a phospholipid bilayer. A Palmitoyl group was then covalently linked to C502 and C503 sulphur atoms to form the single and double palmitoylated calnexin models.

All simulations were performed using the NAMD (Phillips *et al*, 2005) engine, in combination with the CHARMM27 force field brooks, including CMAP corrections. TIP3P water (Jorgensen *et al*, 1983) parameterization was used to describe the water molecules. The spatial overlapping of lipid molecules and protein were removed and the resulting protein-membrane system was solvated in variable-size water box, at a salt concentration of 150 mM NaCl. The periodic electrostatic interactions were computed using the particle-mesh Ewald summation with a grid spacing smaller than 1 Å. All systems were first minimized by 2000 conjugate gradient steps, and subsequently gradually heated from 0 to 300 K in 800 ps with a constraint on the protein backbone scaffold. Finally, the systems were equilibrated for 10 ns at 300 K. Free MD of all equilibrated system were run for not less than 50 ns with a 2-fs integration time step using the RATTLE algorithm applied to all bonds, and trajectories collected for analysis. Constant temperature (300 K) was imposed by using Langevin dynamics (Brunger and Brooks, 1984), with damping coefficient of 1.0 ps. Constant pressure of 1 a.t.m. was maintained with a Langevin piston dynamics (Feller *et al*, 1995) 200 fs decay period and 50 fs time constant.

References

- Abrami L, Bischofberger M, Kunz B, Groux R, van der Goot FG (2010) Endocytosis of the anthrax toxin is mediated by clathrin, actin and unconventional adaptors. *PLoS Pathog* **6**: e1000792
- Abrami L, Kunz B, Iacovache I, van der Goot FG (2008) Palmitoylation and ubiquitination regulate exit of the Wnt signaling protein LRP6 from the endoplasmic reticulum. *Proc Natl Acad Sci USA* **105**: 5384–5389
- Abrami L, Leppla SH, van der Goot FG (2006) Receptor palmitoylation and ubiquitination regulate anthrax toxin endocytosis. *J Cell Biol* **172**: 309–320
- Aebi M, Bernasconi R, Clerc S, Molinari M (2010) N-glycan structures: recognition and processing in the ER. *Trends Biochem Sci* **35**: 74–82
- Alvarez E, Girones N, Davis RJ (1990) A point mutation in the cytoplasmic domain of the transferrin receptor inhibits endocytosis. *Biochem J* **267**: 31–35
- Braakman I, Bulleid NJ (2011) Protein folding and modification in the mammalian endoplasmic reticulum. *Annu Rev Biochem* **80**: 71–99
- Brunger A, Brooks CL (1984) Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem Phys Lett* **105**: 495–500
- Chang DK, Cheng SF, Trivedi VD, Lin KL (1999) Proline affects oligomerization of a coiled coil by inducing a kink in a long helix. *J Struct Biol* **128**: 270–279
- Charollais J, Van Der Goot FG (2009) Palmitoylation of membrane proteins (review). *Mol Membr Biol* **26**: 55–66
- Chavan M, Lennarz W (2006) The molecular basis of coupling of translocation and N-glycosylation. *Trends Biochem Sci* **31**: 17–20
- Chen W, Helenius J, Braakman I, Helenius A (1995) Cotranslational folding and calnexin binding during glycoprotein synthesis. *Proc Natl Acad Sci USA* **92**: 6229–6233
- Chevet E, Smirle J, Cameron PH, Thomas DY, Bergeron JJ (2010) Calnexin phosphorylation: linking cytoplasmic signalling to endoplasmic reticulum lumenal functions. *Semin Cell Dev Biol* **21**: 486–490
- Chevet E, Wong HN, Gerber D, Cochet C, Fazel A, Cameron PH, Gushue JN, Thomas DY, Bergeron JJ (1999) Phosphorylation by CK2 and MAPK enhances calnexin association with ribosomes. *EMBO J* **18**: 3655–3666
- de Brito OM, Scorrano L (2010) An intimate liaison: spatial organization of the endoplasmic reticulum-mitochondria relationship. *EMBO J* **29**: 2715–2723

Statistical analysis

All experiments were performed at least three times, independently. Two-tailed *t*-tests were performed to evaluate the significance of the data.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

Acknowledgements

We thank L Symul for the graphics work in Figure 8, G Davidson for the myc-LRP6 plasmid, C Zurzolo for the PrP plasmid, K Strub for the PPL-3f plasmid and anti-L12 antibodies, R Hedge for anti-TRAP α antibodies. This work was supported by the Swiss National Science Foundation and the EuroMembrane in the context of the EUROCORES program from the European Science Foundation.

Author contributions: AKKL and LA conceived and performed the experiments, analysed the data and wrote the manuscript. TL and MDP performed the molecular dynamics experiments and critically read the manuscript. SB and BK performed the experiments. AK provided the DHHC-expressing plasmids. FGVDG conceived the experiments, analysed the data and wrote the manuscript.

Conflict of interest

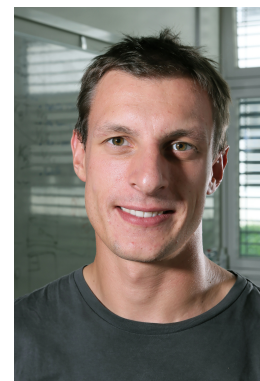
The authors declare that they have no conflict of interest.

- Dekker FJ, Rocks O, Vartak N, Menninger S, Hedberg C, Balamurugan R, Wetzel S, Renner S, Gerauer M, Scholermann B, Rusch M, Kramer JW, Rauh D, Coates GW, Brunsfeld L, Bastiaens PI, Waldmann H (2010) Small-molecule inhibition of APT1 affects Ras localization and signaling. *Nat Chem Biol* **6**: 449–456
- Deprez P, Gautschi M, Helenius A (2005) More than one glycan is needed for ER glucosidase II to allow entry of glycoproteins into the calnexin/calreticulin cycle. *Mol Cell* **19**: 183–195
- Du Y, Ferro-Novick S, Novick P (2004) Dynamics and inheritance of the endoplasmic reticulum. *J Cell Sci* **117**: 2871–2878
- Feller SE, Zhang Y, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation: the Langevin piston method. *J Chem Phys* **103**: 4613
- Fivaz M, Vilbois F, Pasquali C, van der Goot FG (2000) Analysis of GPI-anchored proteins by two-dimensional gel electrophoresis. *Electrophoresis* **21**: 3351–3356
- Fons RD, Bogert BA, Hegde RS (2003) Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. *J Cell Biol* **160**: 529–539
- Gorleku OA, Barns AM, Prescott GR, Greaves J, Chamberlain LH (2011) Endoplasmic reticulum localization of DHHC palmitoyl transferases mediated by lysine-based sorting signals. *J Biol Chem* **286**: 39573–39584
- Greaves J, Chamberlain LH (2011) DHHC palmitoyl transferases: substrate interactions and (patho)physiology. *Trends Biochem Sci* **36**: 245–253
- Hammond C, Helenius A (1994) Folding of VSV G protein: sequential interaction with BiP and calnexin. *Science* **266**: 456–458
- Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* **14**: 33–38, 27–38
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**: 926
- Kordyukova LV, Serebryakova MV, Baratova LA, Veit M (2010) Site-specific attachment of palmitate or stearate to cytoplasmic versus transmembrane cysteines is a common feature of viral spike proteins. *Virology* **398**: 49–56
- Lakkaraju AK, Mary C, Scherrer A, Johnson AE, Strub K (2008) SRP keeps polypeptides translocation-competent by slowing translation to match limiting ER-targeting sites. *Cell* **133**: 440–451

- Lee AH, Iwakoshi NN, Glimcher LH (2003) XBP-1 regulates a subset of endoplasmic reticulum resident chaperone genes in the unfolded protein response. *Mol Cell Biol* **23**: 7448–7459
- Levental I, Lingwood D, Grzybek M, Coskun U, Simons K (2010) Palmitoylation regulates raft affinity for the majority of integral raft proteins. *Proc Natl Acad Sci USA* **107**: 22050–22054
- Li H, Chavan M, Schindelin H, Lennarz WJ (2008) Structure of the oligosaccharyl transferase complex at 12 Å resolution. *Structure* **16**: 432–440
- Linder ME, Deschenes RJ (2007) Palmitoylation: policing protein stability and traffic. *Nat Rev Mol Cell Biol* **8**: 74–84
- Lombardi ML, Jaalouk DE, Shanahan CM, Burke B, Roux KJ, Lammerding J (2011) The interaction between nesprins and sun proteins at the nuclear envelope is critical for force transmission between the nucleus and cytoskeleton. *J Biol Chem* **286**: 26743–26753
- Lynes EM, Bui M, Yap MC, Benson MD, Schneider B, Ellgaard L, Berthiaume LG, Simmen T (2011) Palmitoylated TMX and calnexin target to the mitochondria-associated membrane. *EMBO J* **31**: 457–470
- Martin BR, Cravatt BF (2009) Large-scale profiling of protein palmitoylation in mammalian cells. *Nat Methods* **6**: 135–138
- McGinnes LW, Morrison TG (1994) The role of the individual cysteine residues in the formation of the mature, antigenic HN protein of Newcastle disease virus. *Virology* **200**: 470–483
- Menetret JF, Hegde RS, Aguiar M, Gygi SP, Park E, Rapoport TA, Akey CW (2008) Single copies of Sec61 and TRAP associate with a nontranslating mammalian ribosome. *Structure* **16**: 1126–1137
- Merrick BA, Dhungana S, Williams JG, Aloor JJ, Peddada S, Tomer KB, Fessler MB (2011) Proteomic profiling of S-acylated macrophage proteins identifies a role for palmitoylation in mitochondrial targeting of phospholipid scramblase 3. *Mol Cell Proteomics* **10**: M110.006007
- Myhill N, Lynes EM, Nanji JA, Blagoveshchenskaya AD, Fei H, Carmine Simmen K, Cooper TJ, Thomas G, Simmen T (2008) The subcellular distribution of calnexin is mediated by PACS-2. *Mol Biol Cell* **19**: 2777–2788
- Nagaya H, Tamura T, Higa-Nishiyama A, Ohashi K, Takeuchi M, Hashimoto H, Hatsuzawa K, Kinjo M, Okada T, Wada I (2008) Regulated motion of glycoproteins revealed by direct visualization of a single cargo in the endoplasmic reticulum. *J Cell Biol* **180**: 129–143
- Ohno Y, Kihara A, Sano T, Igarashi Y (2006) Intracellular localization and tissue-specific distribution of human and yeast DHHC cysteine-rich domain-containing proteins. *Biochim Biophys Acta* **1761**: 474–483
- Pendin D, McNew JA, Daga A (2011) Balancing ER dynamics: shaping, bending, severing, and mending membranes. *Curr Opin Cell Biol* **4**: 435–442
- Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* **26**: 1781–1802
- Puhka M, Vihinen H, Joensuu M, Jokitalo E (2007) Endoplasmic reticulum remains continuous and undergoes sheet-to-tubule transformation during cell division in mammalian cells. *J Cell Biol* **179**: 895–909
- Rocks O, Gerauer M, Vartak N, Koch S, Huang ZP, Pechlivanis M, Kuhlmann J, Brunsvelde L, Chandra A, Ellinger B, Waldmann H, Bastiaens PI (2010) The palmitoylation machinery is a spatially organizing system for peripheral membrane proteins. *Cell* **141**: 458–471
- Shibata Y, Shemesh T, Prinz WA, Palazzo AF, Kozlov MM, Rapoport TA (2010) Mechanisms determining the morphology of the peripheral ER. *Cell* **143**: 774–788
- Shibata Y, Voss C, Rist JM, Hu J, Rapoport TA, Prinz WA, Voeltz GK (2008) The reticulon and DP1/Yop1p proteins form immobile oligomers in the tubular endoplasmic reticulum. *J Biol Chem* **283**: 18892–18904
- Skach WR (2007) The expanding role of the ER translocon in membrane protein folding. *J Cell Biol* **179**: 1333–1335
- Viklund H, Bernsel A, Skwark M, Elofsson A (2008) SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. *Bioinformatics* **24**: 2928–2929
- Yang W, Di Vizio D, Kirchner M, Steen H, Freeman MR (2010) Proteome scale characterization of human S-acylated proteins in lipid raft-enriched and non-raft membranes. *Mol Cell Proteomics* **9**: 54–70
- Yount JS, Moltedo B, Yang YY, Charron G, Moran TM, Lopez CB, Hang HC (2010) Palmitoylome profiling reveals S-palmitoylation-dependent antiviral activity of IFITM3. *Nat Chem Biol* **6**: 610–614
- Zuckerman DM, Hicks SW, Charron G, Hang HC, Machamer CE (2011) Differential regulation of two palmitoylation sites in the cytoplasmic tail of the β 1-adrenergic receptor. *J Biol Chem* **286**: 19014–19023

THOMAS LEMMIN

Date of birth: 31.08.1983
Nationality: Swiss/USA
Residence: Chemin des Triaudes 12, 1024 Ecublens,
Switzerland
Languages: French (mother tongue), English, German
Mobile: +4177 421 75 09
E-mail: thomas.lemmin@epfl.ch



ACADEMIC BACKGROUND

- 2013** - **Doctoral Student**, Institute of Bioengineering, Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, Supervisor: Prof. Matteo Dal Peraro
- 2007** - **MSc** in Electrical Engineering and Electronics, Information Technology, EPFL, Switzerland
- Master's Thesis: *Motion artifacts in fMRI*, Advanced Telecommunication Research Institute International (ATR), Japan
 - Minor in Biomedical Engineering, EPFL, Switzerland
- 2005** - **BSc** in Electrical Engineering and Electronics, EPFL, Switzerland

POSITIONS OF RESPONSIBILITY

- Current**
- Teaching assistant for *Bioinformatics* and *Biomolecular Structure and Mechanics*, EPFL
 - Supervision of student projects, Laboratory for Biomolecular Modeling, EPFL
 - Cluster manager at Laboratory for Biomolecular Modeling, (48 nodes), EPFL
 - Workstation manager at Laboratory for Biomolecular Modeling, (20 workstations), EPFL
- Previous**
- Research assistant in neuroradiology, University of Zurich, Switzerland, 2007-2008
 - Teaching assistant for C/C++ programming, EPFL, 2006-2007
 - Electrical draftsman, Energie Ouest Suisse (EOS), Lausanne, Switzerland, 2005-2006.
 - Leader at the YMCA (1990-2001), Federal Certificate (Jeunesse et Sport)

COMPUTATIONAL SKILLS

Operating Systems: Linux (Redhat, Debian), Mac OS X
Software: Matlab, MS Office, LaTeX, Adobe Illustrator, Gimp
Molecular Modeling: NAMD, Gromacs, LAMMPS, Modeller, VMD, Hex
Programming/Scripting: Shell, Tcl, C/C++, Java, Perl

PERSONAL INTERESTS

Rowing • trekking • cross-country skiing • bicycling • Taekwondo

PUBLICATIONS

Lemmin T., Dimitrov M., Fraering P. C., and Dal Peraro M., *The amyloid precursor protein APP-TM is highly flexible and adapts to the lipid environment*, 2012, submitted

Dimitrov M., Alattia J., Lemmin T., Rajwinder L., Fligier A., Houacine J., Hussain I., Radtke F., Dal Peraro M., Behr D., and Fraering P. C., *Early-onset Alzheimer's disease mutations in APP, but not γ -secretase modulators, affect epsilon-cleavage-dependent AICD production*, 2013, submitted

Yan S., Lemmin T., Salvi S., Lausch E., Superti-Furga A., Dal Peraro M., and van der Goot F. G., *Differential effects of frameshift mutations in the exon13 hotspot leading to Hyaline Fibromatosis Syndrome*, under review

Lemmin T., Bovigny C., Lançon D., and Dal Peraro M., *Cardiolipin models for molecular simulations of bacterial and mitochondrial membranes*, Journal of Chemical Theory and Computation, 2012

Lemmin T., Soto C. S., Clinthorne G., DeGrado W. F., and Dal Peraro M., *Assembly of the transmembrane domain of E. coli PhoQ histidine kinase: implications for signal transduction from molecular simulations*, PloS Computational Biology, 2012

Lakkaraju A.K.K., Abrami L., Lemmin T., Blaskovic S., Kunz B., Kihara A., Dal Peraro M., and van der Goot F. G., *Palmitoylated calnexin is a key component of the ribosome-translocon complex*, EMBO Journal, 2012

Lemmin T., Ganesh G., Gassert R., Burdet E., Kawato M., and Haruno M., *Model-based attenuation of movement artifacts in fMRI*, Journal of Neuroscience Methods, 2010

CONFERENCES AND SEMINARS

Biophysical Society, Philadelphia, USA, 2013

A Study of Amyloid Precursor Protein transmembrane domain (poster)

CADMOS: Workshop On Large Scale Computing, EPFL, Switzerland, 2012

Exploring the bacterial membrane (presentation)

Biophysical Society, San Diego, USA, 2012

Modeling of cardiolipin-rich membrane (poster)

WATOC Ninth Triennial Congress, Santiago de Compostela, Spain, 2011

Insights into the signal transduction mechanism of E. Coli PhoQ (poster)

FEBS Congress, Gothenburg, Sweden, 2010

Characterization of the assembly of the PhoQ transmembrane domain (poster)

CSC Workshop, *International Workshop on Coarse-Grained Biomolecular Modeling*, Levi, Finland, 2010

TCBG Workshop, *Computation Biophysics*, Berckman Institute, Urbana-Champaign, USA, 2009

CECAM Workshop, *Methods in Molecular Simulation*, University of Sheffield, UK, 2009